The Report Committee for Lei Wang
Certifies that this is the approved version of the following report:

Patient Flow Optimization at Seton Healthcare

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

David Morton

John Hasenbein

Patient Flow Optimization at Seton Healthcare

by

Lei Wang, B.S.; M.S.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

August 2011

Dedication

To my family for their love and support.

Acknowledgements

First and foremost, I am deeply indebted to my supervisor, Professor David Morton, for his invaluable guidance, insight and support throughout my study, without which this project would be impossible.

I am very grateful to Dr. Jennifer Duthie, for bringing me into this fascinating project, and providing valuable feedbacks during this work.

I wish to thank the people of Seton clinics, especially Jim Thomas, for gratefully sharing their experience and data.

Abstract

Patient Flow Optimization at Seton Healthcare

Lei Wang, M.S.E

The University of Texas at Austin, 2011

Supervisor: David Morton

We analyze the patient flow of three community health clinics from the Seton group in Austin, Texas, using simulation tools. Our goal is to help the clinics find solutions to cope with increasing patient demand. Several scenarios for increasing efficiency are explored using an ARENA-based patient flow model. Multiple bottlenecks are identified and solutions are found to help the clinics minimize overall patient cycle time and to distribute the workload more evenly across the staff. This study demonstrates that healthcare service facilities may benefit from quantitative analysis, especially simulation tools, to improve their efficiency.

Table of Contents

List of Tables

List of Figures

List of Illustrations

Chapter 1 Problem statement

1.1 BACKGROUND

1.1.1 Healthcare industry background

Healthcare service consumes a growing fraction of the economic production, in the United States and also other countries in the Western world. This increasing cost is due to aging populations as well as the introduction of expensive new treatments. By the end of this decade, it is forecasted that the registered nurse (RN) workforce will fall approximately 12 percent below requirements. Indeed, the U.S. Department of Health and Human Services predicts that by 2020, the nurse shortage will reach between 350,000 and 800,000 [1]. Due to the aging of the post-World War II baby-boom generation, such a shortage must be treated seriously. Therefore, the administrators of healthcare facilities must deliver quality healthcare via appropriate resource allocation.

Patient flow is critical to the quality of healthcare delivery: A more predictable patient flow can lead to both better staffing and better care. Specifically, patient flow is the ability of a healthcare facility to serve patients promptly and efficiently when moving through stages of healthcare service [2]. That is, good patient flow means the patients experience minimal queueing delay. Understanding patient flows can help healthcare facilities to improve their operations, including resource planning, scheduling, and utilization.

The second reason to investigate the patient flow is due to the recent changes in healthcare reimbursement. In the mid-1980s, the prospective payment system (PPS) was adopted by the federal government, which pays healthcare providers fixed amount for their services based on the patient’s diagnosis. This gives the healthcare providers incentive to improve their efficiency in the delivery of services. Consequently, they started to concentrate on understanding and optimizing their patient flows [3].

1.1.2 About Seton

The Seton Family of Hospitals is among the most important health organizations in Texas, serving an 11-county area with a population of 1.8 million [4]. The organization operates the following healthcare facilities:

- Five major medical centers,
- two community hospitals,
- two rural hospitals,
- an inpatient mental health hospital, and
- three primary care clinics (McCarthy, Kozmetsky and Topfer).

The three primary care clinics (McCarthy, Kozmetsky and Topfer) provide affordable healthcare services, by primary care physicians, nurse practitioners and registered nurses, to medically underserved Austin residents [4]. In the project leading to this report, we worked with the staff of these clinics to understand and optimize their current operations.

1.2 Literature review

Like many other community clinics, Seton clinics face the problem of increasing demand but have limited resources to meet that demand. In this section we review research on how waiting time affects patient satisfaction and on how patient waiting time can be reduced.

Huang [5] reports that patient waiting time is often the major cause for a patient's complaints about his/her experience when visiting an outpatient clinic. Huang's results show that generally patients are quite satisfied if the waiting time does not exceed 37 minutes when arriving on time, and does not exceed 63 minutes when the patients are late for appointments. These results can help to set up waiting time limits and to design patient dispatching rules. Based on this, Wilson and Nathan [6] suggest methods to set benchmarks to measure the quality of the healthcare service delivered.

In the rest of this section, we summarize the literature in three related areas: general research and guidelines on how to reduce waiting time in clinics and physicians' offices, research on how to improve patient flow based on statistical tools and simulation, and research on the relationship between waiting time and appointment intervals.

1.2.1 General guides on patient flow improvement

Nolan et al. [7] give detailed guidelines on how to reduce delay and waiting time. The authors identify four main causes for delay in clinics and physicians' offices, namely, schedules are overbooked, processes are not synchronized, physicians are unavailable, and demand for urgent care can vary. Nolan et al. discuss corresponding solutions for each of these causes.

Langley et al. [8] offer an integrated approach to process improvement of quality and productivity in diverse settings. Their approach has been used successfully by hundreds of healthcare organizations to improve healthcare processes and outcomes. The approach has two parts:

First, there are three fundamental questions to answer:

- What are our goals?
- How to ensure that a change is helpful?
- What changes will result in an improvement?

Second, to assess whether a change is an improvement, use a Plan-Do-Study-Act cycle to test and implement changes in real work settings.

1.2.2 Research based on statistical tools

In this section we review work on how to improve patient flow using statistical and simulation tools. Racine and Davidson [9] investigate an academic pediatric practice, and study changes in practice patterns and the consequent outcomes in performance measurement including patient waiting times, total visit times, and room and nursing time usage rates, based on data from a two-week time-flow study (one week in year 1999 and one week in year 2000). Specifically, the authors analyze year-1999 data from the time-flow of the first week, and propose several areas for potential improvements, including (1) nurse usage is not optimal when nurses are required to do administrative tasks including answering telephone calls; (2) practitioners are not evenly distributed, resulting in over-crowding of examination rooms on some days, and under-use on other days; (3) nurses are centralized to two stations, which make it difficult for the nurses to monitor and respond to patient flow needs; and, (4) there is excessive time demand for documentation. Then the authors recommend solutions for the above-mentioned issues.

After those solutions were implemented, in year 2000, data was collected again for one week, to compare with the initial data. Racine and Davidson perform a statistical analysis and show convincingly that time-flow studies can be useful tools to identify and mitigate inefficiency in the delivery of healthcare service. Potisek et al. [10] conduct a similar study except that the focus is on patients with chronic conditions.

1.2.3 Research based on simulation on patient flow improvement

Besides statistical methods, discrete-event simulation is another widely used analysis tool in industry settings that are often highly complicated and uncertain [11]. For example, Merkle [12] builds simulation models to analyze the process in the family clinics of The Brooke Army Medical Center, and to support decision making based on sensitivity analysis of the model. Similarly, Jun et al. [14] design a discrete-event simulation model of a physician clinic environment. Rohleder et al. [15] also describe a similar study on using simulation methods to redesign phlebotomy and specimen collection centers at a medical diagnostic laboratory. The authors show how a system dynamics model can help predict and address implementation problems in healthcare facilities. For further references to further related work, see the survey article of Swisher et al. [13].

1.2.4 Research on correlation between patients' waiting time and appointment intervals

We review work on analyzing the correlation between patient waiting time and the way appointments are reserved. Hill-Smith [16] reports that the expected waiting time of a patient increases exponentially as the interval of the appointment is reduced. Hill-Smith finds that appointment intervals less than the median length of a consultation can be counter-productive. Dexter [17] suggests that when all appointment slots are assigned to other patients, an add-on patient should either be seen by a different provider or at the end of the regular clinic session. He also suggests measuring the average consultation times accurately for each provider, so that computer simulation can be performed using parameters appropriate for each specific provider.

Chapter 2 Data collection and analysis

2.1 Typical path of patient flow in a Seton clinic site

2.1.1 Operations in a Seton clinic site

Our study involves three Seton primary clinic practices located in northern (Topfer), central (McCarthy) and southern (Kozmetsky) communities in Austin, Texas. Each clinic serves patients from 8:30am to 6pm from Monday to Thursday and from 8:45am to 4:30pm on Friday. There are two types of days in the clinical operation: a 2-provider day and a 3-provider day. At the McCarthy site, it is a 2-provider day, except when a certain specialist is scheduled in the clinic, when it is instead then a 3-provider day. At the Topfer site, three days per week it is a 3-provider day and the other two days it is a 2-provider day. At the Kozmetsky site, there are 1.5 days each week when there are three providers serving in the clinic and during the other 3.5 days there are two providers serving in the clinic. Each provider, i.e., physician, is teamed with a clinical assistant (CA) and a registered nurse (RN) to serve the patient. Patients are scheduled every 15 minutes from 8:45am to 11:30am in the morning and from 1:45pm to 4:30pm in the afternoon for each provider.

The next two sections briefly describe the path of two types of patients, one with a provider appointment and one with a nurse appointment. In addition, our simulation model includes two more patient types, walk-in patients and those whose visit is handled by a social worker. These latter two types of patients have very simple paths through the clinic. We discuss them in further detail in Chapter 3.

2.1.2 Typical path of a patient with provider appointment

The patient flow begins with patient check-in at the front desk by a customer service representative (CSR). Clinic assistants (CAs) are then notified of the patient's arrival and a CA then brings the patient to an exam room. Here, the CA performs a basic patient assessment, which consists of collecting the patient's height, weight, temperature, etc. After this, the patient waits in the exam room and the CA flags the room to notify a provider that the patient is ready for examination. The patient is then seen by the provider. Upon completion, the provider exits the exam room, writes up a treatment plan and flags the patient for discharge. A registered nurse (RN) then administers any vaccinations or treatment noted by the provider. Once discharge is complete, the patient either leaves the clinic or returns to the front desk to schedule another appointment with the provider.

2.1.3 Typical path of a patient with nurse appointment

The patient flow begins with patient check-in at the front desk by a CSR. The RNs are then notified of the patient's arrival and an RN then brings the patient to an exam room. When the exam is finished, the RN discharges the patient. Once discharge is complete, the patient either leaves the clinic or returns to the front desk to schedule another appointment with an RN.

2.2 DATA DESCRIPTION

2.2.1 Initial data collection

Based on the above-described flow of patients, a collaborative effort was undertaken to collect two weeks of data on patient sojourn times at the McCarthy clinic in December 2010, the Topfer clinic in March 2011, and the Kozmetsky clinic in May 2011. During the two weeks in each clinical site, 428 data points were collected from Topfer, 433 data points from McCarthy and 430 data points from Kozmetsky. Every CSR, CA, RN, and provider recorded both the the beginning and the end of his/her activity with a patient at each stage of the visit. This allowed us to monitor every visit during the two weeks as well as the engagement of the staff. The recording sheet we used, as shown in Illustration 1, is adapted from Racine and Davidson [9], with revisions of the description of each stage to match the process of Seton clinics.

Patient Flow Study

Date:

Patient's Name:

Provider:

Type of visit:

		TIME	COMMENTS
CSR	1.Appointment		
	2.Arrival		
	3.Check-out (Other type appt)		
CA	4.Call in by CA		Hearing/Vision Y N Labs? Y N
	5.CA vital exam complete		
Provider	6.Provider starts		
	7.Provider ends		
RN	8.Nurse retrieve the chart		
	9.Discharge starts		Immunizations? Y N Labs? Y N
	10.Discharge ends		

Illustration 1: Patient flow study form

For instance, the CSR recorded a patient's arrival time and the time of his/her appointment. CAs recorded when they called a patient into an exam room, and the beginning and end times of assessing the patient's vital signs. The providers recorded the time when the exam began and was completed. The RN recorded when charts were retrieved from the provider to begin discharging procedures and when the discharge finished. Intervals between the end of one patient's activity and the beginning of the next were used to estimate waiting time of a patient between different stages. The beginning time of the visit is set as the actual arrival time of the patient regardless of whether the patient arrived early or late from his/her appointment time. The total time a patient spends in the clinic is calculated by subtracting the arrival time from the discharge time.

2.2.2 Description of data in each stage of patient flow

All the data collected can be categorized into two segments: input data and output data. Input data are primitive service times and patient's arrival times while output data are performance measures that will be also collected from the simulation model.

Input data

Distributions of process variables

- Deviation of arrival time from appointment time
- Inter-arrival time of patients
- Service time for vital assessment
- Service time for provider's exam
- Service time for nurse's discharge

Number of resources

- Number of CAs
- Number of providers
- Number of nurses
- Number of exam rooms

Output data

Waiting time of patients in each stage

- Waiting time for CA to call for vital assessment
- Waiting time for provider's exam
- Waiting time for discharge
- Total cycle time

We consider two possible models of the arrival process of patients. One model treats inter-arrival times as i.i.d. random variables, while ignoring appointment times of the patients. As we will see in the later chapters, the simulation results, particularly total cycle time, based on this model do not fit the collected data well. An alternative model assumes that the arrival time of a patient depends upon his/her scheduled appointment time. As such, this model treats the deviation of the arrival from appointment time as i.i.d. random variables. We discuss both models in this chapter.

2.2.2.1 Deviation of arrival from appointment time

In this section we discuss statistics of customers whose arrival time deviates from their appointments. We model separately early deviation and late deviation. Statistics of early deviation, i.e., arrival times earlier than the appointment time are depicted in Table 1. As depicted in Table 1, when patients are early, they arrive the clinic on average 19 to 22 minutes before their schedule. While when patients are late, they arrive the clinic on average nine to ten minutes after their scheduled appointment time, as shown in Table 2.

Time	Site	Mean	Standard Deviation	Minimum Value	Maximum Value
Deviation of arrival time from appointment time (early)	McCarthy	19.1	19.2	1	102
	Kozmetsky	21.9	19.5	1	120
	Topfer	19.5	19.5	1	133

Table 1: Summary statistics of early deviation of arrival from appointment time (minutes)

Time	Site	Mean	Standard Deviation	Minimum Value	Maximum Value
Deviation of arrival time from appointment time (late)	McCarthy	9.47	8.05	1	35
	Kozmetsky	10.1	10.4	1	67
	Topfer	10.5	11.4	1	60

Table 2: Summary statistics of late deviation of arrival from appointment time (minutes)

2.2.2.2 Inter-arrival time

Summary statistics of patient inter-arrival times for each clinic are depicted in Table 3. The inter-arrival time is calculated by subtracting the patient's arrival time from the arrival time of the next patient. As shown in Table 3, the mean inter-arrival time is around six minutes at the Kozmetsky site and the Topfer site and nine minutes at the McCarthy site.

Time	Site	Mean	Standard Deviation	Minimum Value	Maximum Value
Deviation of arrival time from appointment time (late)	McCarthy	9.47	8.05	1	35
	Kozmetsky	10.1	10.4	1	67
	Topfer	10.5	11.4	1	60

Table 3: Summary statistics of inter-arrival time (minutes)

2.2.2.3 Service time for assessing a patient's vital signs

Summary statistics of the service times to assess a patient's vital signs are depicted in Table 4. The process of vital assessment includes taking the patient's weight and height, temperature, blood pressure and confirming the purpose of the patient's visit. Labs or tests are done for certain types of visits such as well-woman exams, well-child exams and diabetic exams. As shown in Table 4, the average time for CAs at the McCarthy site to take a patient's vital signs is 8.14 minutes while the average time for CAs at the Kozmetsky and the Topfer to process the assessment is 7.66 minutes and 6.83 minutes respectively.

Time	Site	Mean	Standard Deviation	Minimum Value	Maximum Value
Service time for assessing vital signs	McCarthy	8.14	3.15	2	18
	Kozmetsky	7.66	4.38	1	55
	Topfer	6.83	3.97	2	29

Table 4: Summary statistics of service time of assessing vital signs (minutes)

2.2.2.4 Service time for provider's exam

Summary statistics for the duration of a provider's exam are depicted in Table 5. In this stage, a provider exams the patient, addresses the patient's questions and gives prescriptions. The provider updates the charts and gives out lab forms if any lab work needs to be done. We observe from Table 5 that the providers in the McCarthy site need around 22 minutes to examine the patient while the providers in the Kozmetsky site and the Topfer site only need around 13 minutes to complete an examination.

Time	Site	Mean	Standard Deviation	Minimum Value	Maximum Value
Service time for provider's exam	McCarthy	21.8	9.15	5	60
	Kozmetsky	12.9	7.47	3	55
	Topfer	13	7.01	1	47

Table 5: Summary statistics of service time of provider's exam (minutes)

2.2.2.5 Service time for discharge

Summary statistics for the time it takes for a nurse to discharge a patient are depicted in Table 6. In the discharge process, the RN gives instructions to the patient on the issues that need to be taken care of. Vaccination is given and labs are taken if ordered by the provider. As shown in Table 6, it takes around 10.8 minutes for a nurse at McCarthy to discharge a patient and it takes a nurse at Kozmetsky an average of 9.18 minutes to discharge a patient while it takes an average of 7.12 minutes for a nurse at Topfer to process the discharge.

Time	Site	Mean	Standard Deviation	Minimum Value	Maximum Value
Service time for discharge	McCarthy	10.8	8.26	0	54
	Kozmetsky	9.18	7.36	0	42
	Topfer	7.12	6.47	0	30

Table 6: Summary statistics of service time of discharge (minutes)

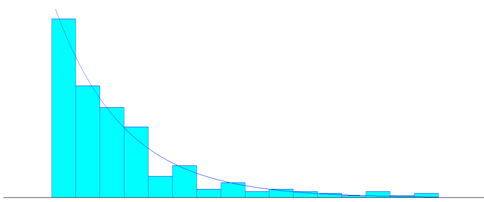
2.3 DATA ANALYSIS

This section is devoted to understanding the input data. More precisely, as we model each input data set as i.i.d. samples from a probability distribution, we use Arena to find both the family of the distributions and the parameters of that distribution that best describe the observed data. We observe that the data collected contains outliers, typically a few samples significantly larger than the rest, which can significantly skew the parameters output by Arena. Hence, we pre-process the data to remove these outlying samples.

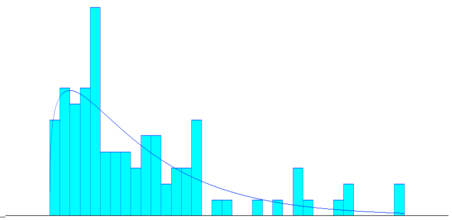
In the following, we use *U(a, b)* to denote a uniform random variable with support (*a, b*), *TRIA(a,b,c)* to denote a triangular distributed random variable with support *(a,c)* and mode *b*, $N(\mu,\sigma^2)$ to denote a normal random variable with mean μ and variance σ^2, $EXPO(\mu)$ to denote an exponential distributed random variable with mean μ, $GAMMA(\theta, k)$ for a gamma random variable with scale parameter θ and shape parameter k, $ERLA(\lambda, k)$ for an Erlang random variable with rate parameter λ and shape parameter k, and $WEIB(\lambda, k)$ for a Weibull random variable with scale parameter λ and shape parameter k.

2.3.1 Deviation of arrival from appointment time

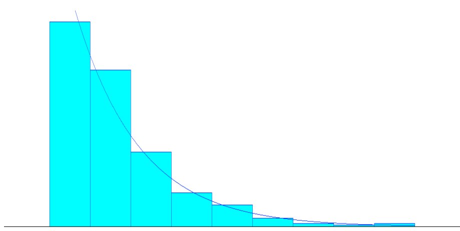
We model the deviation of the arrival time from the appointment time by distinguishing early and late arrivals. In each category, we plug the group of data into Arena's Input Analyzer and obtain the distribution with the best fit, based on a measure of squared error. The results are depicted in Figure 1, along with the goodness-of-fit p-values for the chi-square test.



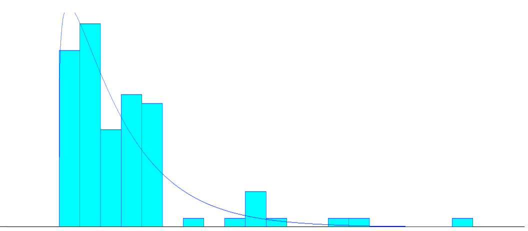
Early (McCarthy)
Histogram Range = 1 – 102 minutes
Number of Intervals =16
Fit-distribution: 0.999+EXPO(18.1)
Chi-square test p-value: 0.175



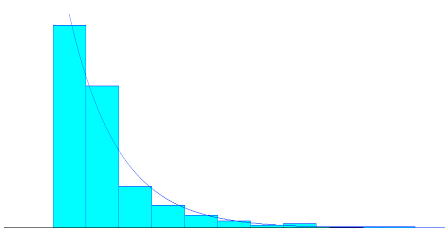
Late (McCarthy)
Histogram Range = 0.5 – 35.5 minutes
Number of Intervals =35
Fit-distribution: 0.5+GAMMA(6.99, 1.28)
Chi-square test p-value: 0.0947



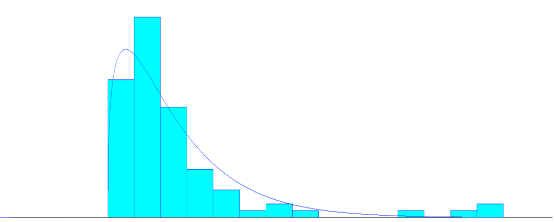
Early (Kozmetsky)
Histogram Range = 1 – 120 minutes
Number of Intervals =9
Fit-distribution: 0.999+EXPO(20.9)
Chi-square test p-value: 0.0696



Late (Kozmetsky)
Histogram Range = 0.5 – 67.5 minutes
Number of Intervals =20
Fit-distribution: 0.5+GAMMA(8.12, 1.19)
Chi-square test p-value: 0.111



Early (Topfer)
Histogram Range = 1 – 150 minutes
Number of Intervals =12
Fit-distribution: 0.999+EXPO(18.5)
Chi-square test p-value: 0.0226

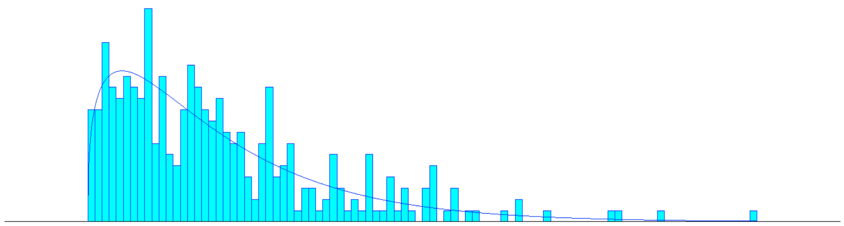


Late (Topfer)
Histogram Range = 0.5 – 60 minutes
Number of Intervals =15
Fit-distribution: 0.5+GAMMA(7.37, 1.36)
Chi-square test p-value: 0.0218

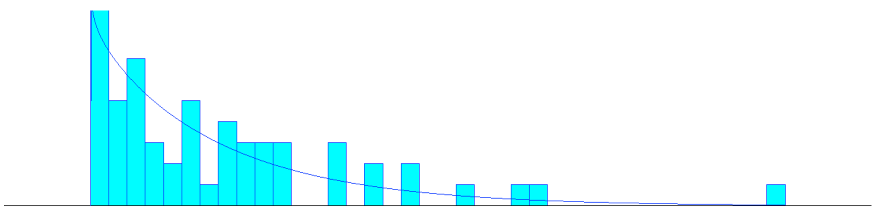
Figure 1: Distribution of deviation of arrival from appointment time

2.3.2 Inter-arrival time

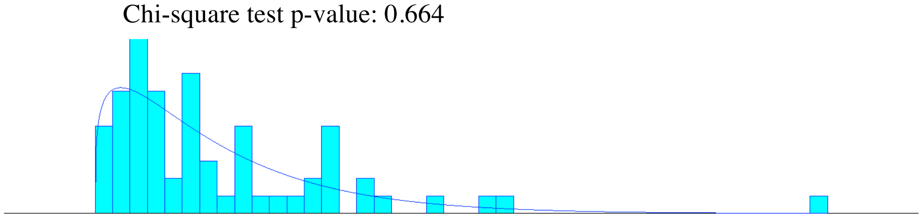
For the McCarthy site, we again employ the Input Analyzer of Arena in order to analyze the data collected for inter-arrival times. The gamma distribution provides a reasonable fit to the data for the McCarthy site with a p-value of 0.14 as shown in Figure 2. We also use the gamma distribution to model inter-arrival times at the Kozmestky and Topfer sites, which are again shown in Figure 2.



McCarthy site
Histogram Range = -0.5 – 93.5 minutes
Number of Intervals =94
Fit-distribution: -0.5+GAMMA(6.52, 1.48)
Chi-square test p-value: 0.14



Kozmestky site
Histogram Range = -0.5 – 37.5 minutes
Number of Intervals =38
Fit-distribution: -0.5+GAMMA(7.48, 0.95)
Chi-square test p-value: 0.664

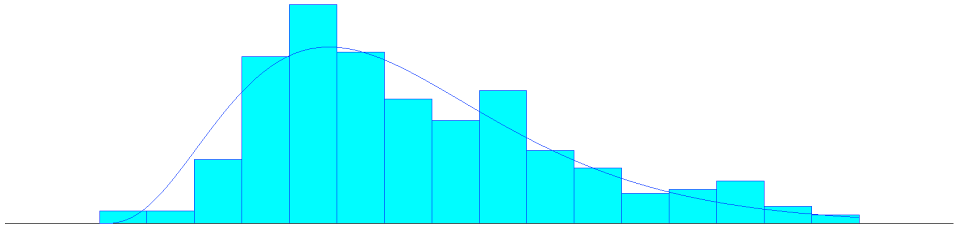


Topfer site
Histogram Range = -0.5 – 41.5 minutes
Number of Intervals =42
Fit-distribution: -0.5+GAMMA(5.53, 1.26)
Chi-square test p-value: 0.0553

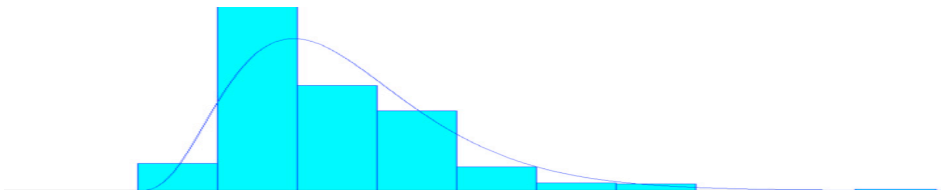
Figure 2: Distribution of inter-arrival time

2.3.3 Service time for assessing a patient's vital signs

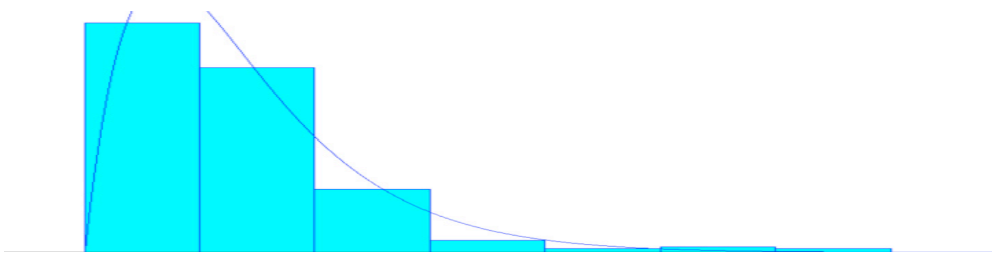
The process of a CA taking a patient's vital signs includes several sub-processes such as taking the patient's weight and height, taking the patient's temperature and blood pressure and confirming the patient's purpose of visit. If the duration of each of these sub-processes is modeled as an exponential random variable then the total service time for the CA to assess a patient's vital sign is naturally modeled using an Erlang distribution. With this motivation, we used Arena's Input Analyzer to fit an Erlang distribution to the data for each of our three clinical sites. The results are depicted in Figure 3.



McCarthy site
Histogram Range = 1.5 – 18 minutes
Number of Intervals =16
Fit-distribution: 1.5+ERLA(1.66, 4)
Chi-square test p-value: 0.0754



Kozmetsky site
Histogram Range = 0.5 – 28 minutes
Number of Intervals =10
Fit-distribution: 0.5+ERLA(1.79, 4)
Chi-square test p-value: <0.005

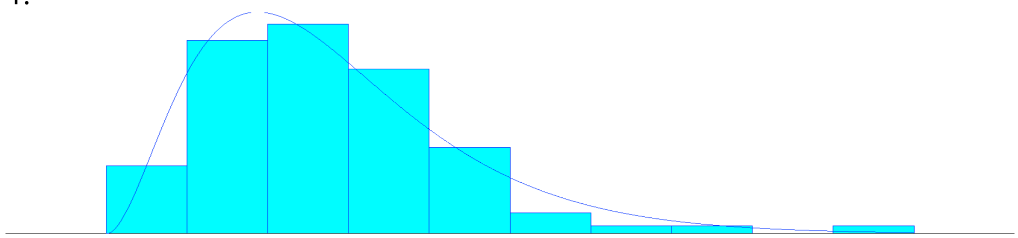


Topfer site
Histogram Range = 1.5 – 29.5 minutes
Number of Intervals =7
Fit-distribution: 1.5+ERLA(2.67, 2)
Chi-square test p-value: 0.0322

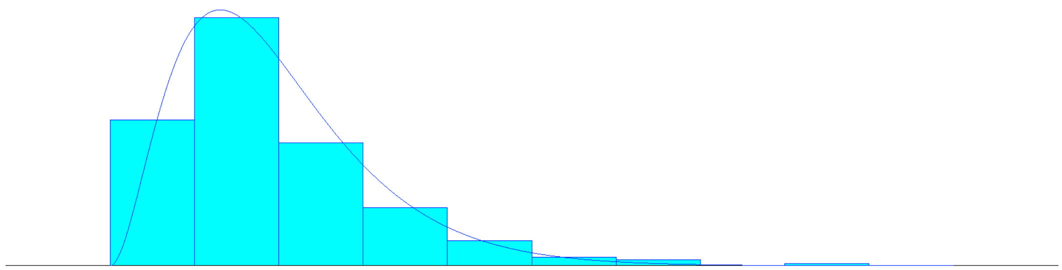
Figure 3: Distribution of service time of assessing vital signs

2.3.4 Service time for provider's exam

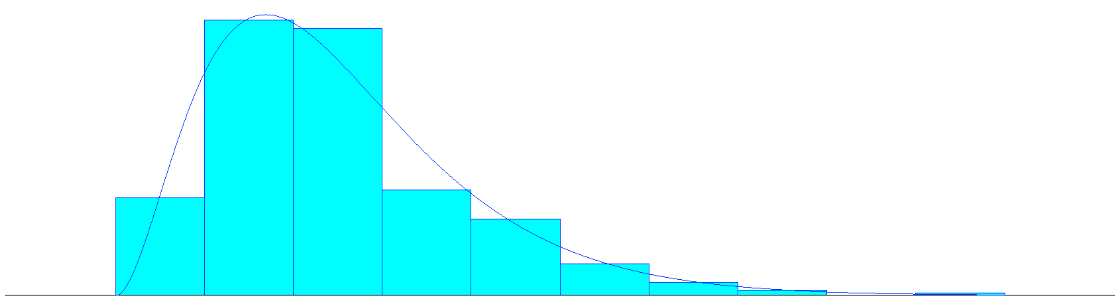
We again apply Arena's Input Analyzer to fit a distribution to the service time for the provider's exam. The process of a provider's exam includes several sub-processes such as examining the patient, updating the chart, etc. For the same reason as above, we choose an Erlang distribution to fit the data. The detail distribution is depicted in Figure 4.



McCarthy site
Histogram Range = 5 – 65 minutes
Number of Intervals =10
Fit-distribution: 5+ERLA(5.62, 3)
Chi-square test p-value: 0.0421



Kozmetsky site
Histogram Range = 1.5 – 60 minutes
Number of Intervals =10
Fit-distribution: 1.5+ERLA(3.81, 3)
Chi-square test p-value: 0.0745

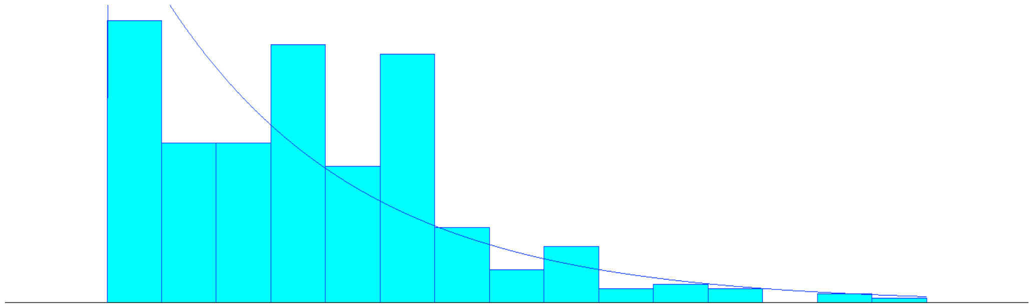


Topfer site
Histogram Range = 0.5 – 50 minutes
Number of Intervals =10
Fit-distribution: 0.5+ERLA(4.18, 3)
Chi-square test p-value: 0.154

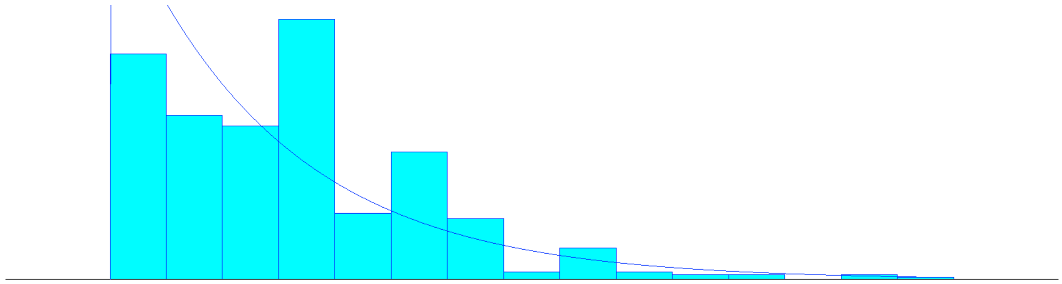
Figure 4: Distribution of service time of provider's exam

2.3.5 Service time for discharge

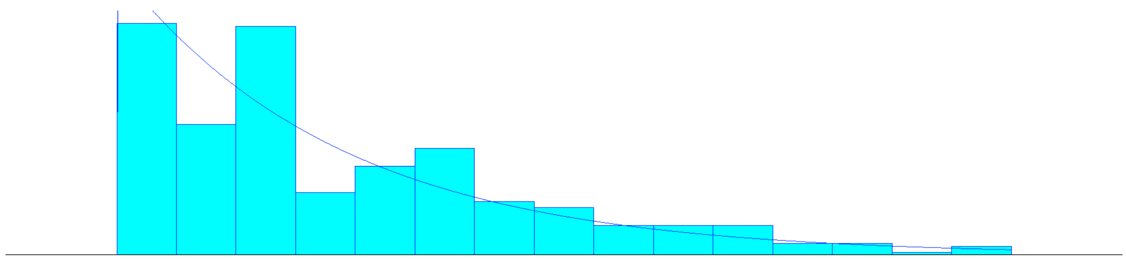
As shown in Figure 5, we also choose the Erlang distribution for the service time for a discharge because the discharge process also includes several sub-processes, for example, giving instructions to the patient, giving vaccines, etc. and it has the smallest mean square error among the fitted distributions. However, from the histograms, it appears that the distribution of discharge time is quite complicated, and cannot be well fitted by distributions supported by Arena. Despite this, we will use these distributions in our simulation model of the next section. Fortunately, the discharge process is at the end of the patient's visit and therefore errors in our ability to model this service time do not have a cascading effect through the rest of the simulation model.



McCarthy site
Histogram Range = 0 – 45 minutes
Number of Intervals =15
Fit-distribution: -0.001+ERLA(10.8, 1)
Chi-square test p-value: <0.005



Kozmetsky site
Histogram Range = -0.5 – 50 minutes
Number of Intervals =15
Fit-distribution: -0.5+ERLA(9.68, 1)
Chi-square test p-value: <0.005



Topfer site
Histogram Range = -0.5 – 30 minutes
Number of Intervals =15
Fit-distribution: -0.5+ERLA(7.62, 1)
Chi-square test p-value: <0.005

Figure 5: Distribution of service time of discharge

Chapter 3 Simulation models

In this chapter we discuss in detail how we build the simulation models. Specifically, we present two simulation models based on our two different approaches to modeling the arrival process of patients. One approach uses i.i.d. inter-arrival times to model the arrival processes, and the other uses the deviation from the appointment time to model the arrival processes.

3.1 Assumptions for the models

We make the following general assumptions on the clinic staff and relevant distributions for tasks they perform and other model primitives, for both models we consider.

General Assumptions:

1. All CAs in the same clinic perform statistically identically. We do not model differences in experience or position level.
2. All providers in the same clinic are statistically identical.
3. All RNs in the same clinic are statistically identical.
4. Patients are identical except in the path that they follow through the clinic. We do not model the differences in age, gender and type of diseases. Patients are distinguished by whether they have a provider appointment or an RN appointment, as described in Chapter 2. In addition, we describe two other types of patients in this chapter.

Assumptions on certain distributions:

In Chapter 2, we discussed the arrival process and service process, which are the major factors in the entire process. Therefore, the distributions associated with these processes are obtained by fitting the collected data with Arena. In this section, we discuss how we model other processes in the system, which are less critical. We use a Poisson process to model the arrival of phone calls and other clerical distractions. This modeling choice is made because each phone call and clerical distraction arrives independently of others, with a somewhat predictable rate of arrivals per hour. The service times for the distractions are modeled with a uniform distribution. This is because the time to complete the tasks is quite variable yet it has a minimum and maximum value. The detailed assumptions on a number of distributions used in the simulation model are listed in Table 7.

Event	Distribution	Reason for Distribution
Check In	U(1,5)	The data vary between 1 and 5 minutes with roughly equal probability based on staff estimation.
Delay Non-provider (such as social workers)	WEIB(33.6,0.794)	From the arrival data we collect when doing on-site observation.
Schedule Appointment	U(2,4)	Min of 2 minutes and a max of 4 minutes with equal probability in between based on staff estimation.
Triage	TRIA(6,8,11)	It usually takes around 8 minutes but can be as low as 6 minutes or as high as 11 minutes based on staff estimation.
Walk-in Patient Helped by CSR	U(3,5)	Min of 3 minutes and a max of 5 minutes with equal probability in between based on staff estimation.
Answer Phone Call	U(2,5)	Staff estimates that the phone calls take anywhere from 2 to 5 minutes with equal probability.
Complete Clerical Task	U(.5,4)	Clerical tasks are quite short and can take anywhere from .5 to 4 minutes.
Nurse Examines the Patient in a nurse appointment	N $(15,5^2)$	Estimated by: 60% of the appointments take 10 to 20 minutes.
Nurse Prepares the Exam Room	U(1,2)	Staff estimates that rooms take no more than 2 minutes to prepare. We estimate that it takes at least 1 minute.

Table 7: Assumptions on other distributions

To construct the simulation model we must also estimate the time it takes the patients to travel around the clinic. We model these "route times" as either constants or uniform distributions, depending on the distance traveled. For instance, the distance from an exam room to the exit is longer than the route from the waiting room to an exam room. So we model some routes with a constant time of one minute and others with a uniform distribution from zero to one minute, as shown in Table 8.

Event	**Distribution**	**Reason for Distribution**
Route to Exam Rooms	U(0,1)	From our observations, we estimate a uniform distribution between 0 and 1 minute.
Route to Exit, Triage and Front Desk for Appointment	Constant 1	From our observations, we estimate a constant route time of 1 minute.

Table 8: Assumptions on distributions for patient routing times

3.2 SIMULATION MODEL WITH INTER-ARRIVAL TIME

In turn, we now describe the simulation model in which we model the arrival process of patients via i.i.d. inter-arrival times and then deviation from appointment times. The next three sections describe the model for the three sites under the former assumption of modeling inter-arrival times. Section 3.3 then does the same using deviation from appointment time.

3.2.1 Model of the McCarthy site

We model the McCarthy site using the inputs listed in Table 9. The distribution of inter-arrival time is listed in Table 9 as -0.5+GAMMA(6.52,1.48). This would allow the possibility of negative inter-arrival times. However, Arena truncates negative realizations to zero.

Input	**Distribution**
Inter-arrival time of patients (minutes)	-0.5+GAMMA(6.52,1.48)
Process time of vital assessment (minutes)	1.5 + ERLA(1.66,4)
Process time of provider's exam (minutes)	5 + ERLA(5.62,3)
Process time of nurse's discharge (minutes)	-0.001+ERLA(10.8,1)
Number of CAs	2
Number of providers	2
Number of nurses	2
Number of exam rooms	8

Table 9: Input data for the McCarthy site (inter-arrival time model)

Our model simulates half of a day at the clinic (from 8:45am to 12:45pm or 240 minutes). We create one version of the model for the McCarthy site with two CAs, two RNs and two providers as shown in Table 9. The main parts of the model are the front desk, assessment, provider's examination, discharge and distractions. We describe the first four parts of the model in Chapter 2 and these involve patients moving through the clinic. The distractions are our method of modeling unexpected occurrences that distract the staff such as phone calls and clerical work. The method of modeling these is described at the end of this section. To begin discussing the logic behind the model, consider the first area: the front desk. Four types of patients enter the system and proceed to the front desk: patients with a provider appointment, patients with a nurse appointment, patients with a non-provider appointment (such as a social worker), and walk-in patients.

We model the arrival process of each of these tour types of patient separately as shown in the screenshot from Arena in Illustration 2. All four patients are then assigned various attributes indicating the patient type. These attributes dictate the patient's path through the clinic in the simulation model. We randomly give patients with a provider appointment a provider index because providers usually schedule the same amount of patients each day and the patients are identical. After the patient arrives at the clinic, he/she is routed from the front entrance to the front desk. Walk-in patients enter the same "Enter Clinic" station as patients with appointments. All patients then either commence service by an available CSR to check in or begin queueing if the CSRs are both busy. After being checked-in, the model ascertains which type of patient is being processed and then routes the patient according to his/her patient type, where he/she is then routed to queue for the appropriate process as shown in Illustration 3.

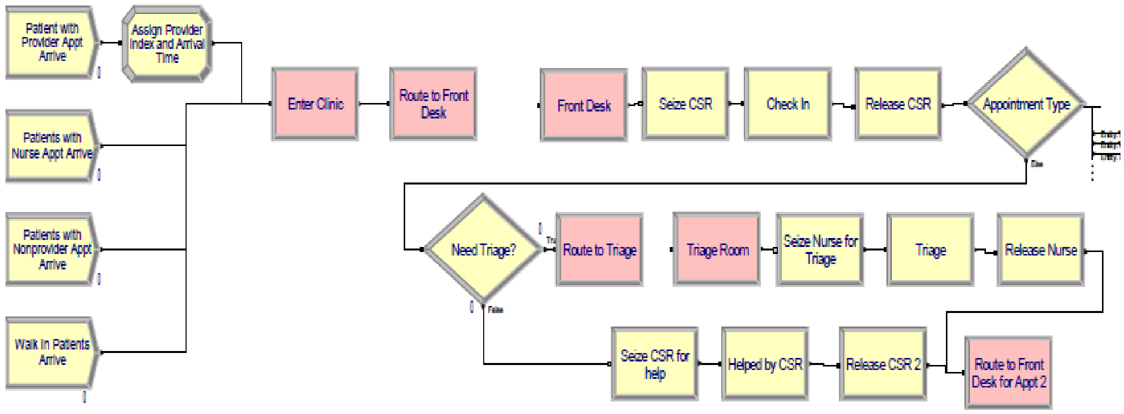


Illustration 2: Illustration of model - patient arrives to the system and checks in

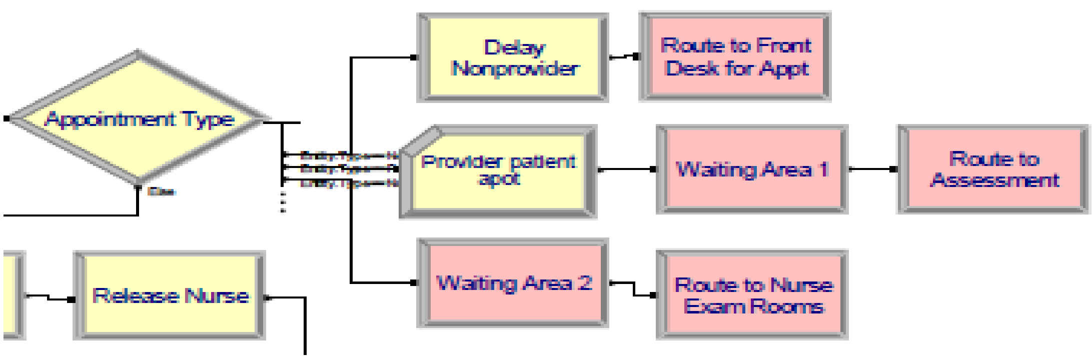


Illustration 3: Illustration of model - patient waits for the next process after check in

Patients with a provider appointment are then routed to the waiting room to queue while waiting for the CA to take vital signs and direct them to an exam room. Patients with a nurse appointment are also routed to the waiting room to queue while waiting for an RN and an exam room (they do not undergo assessment of their vital signs). The patients with a non-provider appointment are not modeled in detail so they are delayed and then they are routed to the CSR to possibly set a future appointment as depicted in Illustration 4.

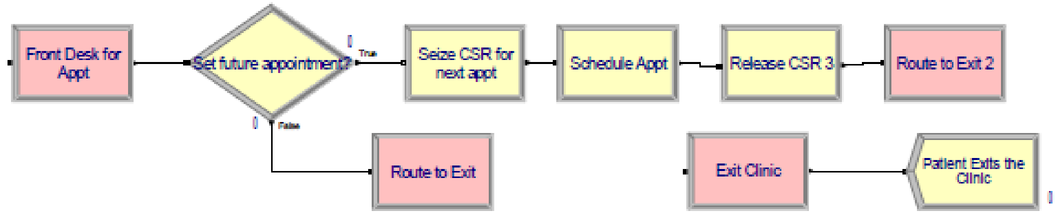


Illustration 4: Illustration of model - patient sets future appointment at the front desk

All patients with an appointment ultimately are routed to the CSR to schedule a follow-up appointment. The last type of patient is a walk-in patient. Such patients either receive service by a CSR (they need simple services such as scheduling an appointment or picking up diabetes supplies) or by an RN (they need triage). With probability 0.25 a walk-in patient requires service by a CSR and with probability 0.75 by a nurse. This fork in a walk-in patient's path is shown in Illustration 5. A patient requiring service by a CSR may have to queue before receiving service and then leaves the system after the service is complete. Patients who need triage proceed to the waiting room where they wait to receive service from an RN. After receiving service from an RN, they visit a CSR to possibly schedule a future appointment prior to exiting the clinic.

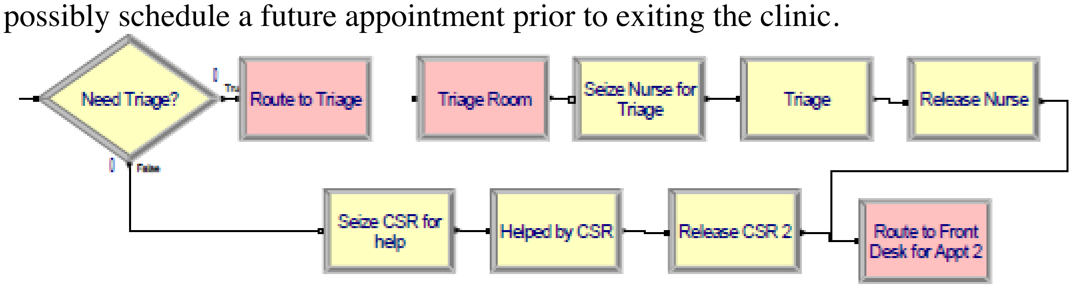


Illustration 5: Illustration of model - the path of a walk-in patient

The next stage for the patients with a provider appointment is assessment of their vital signs. The patients are routed from the waiting room to the exam room station. The provider index of the patient is used to assign the patient to their provider's CA. (We also can assign patients randomly to CAs but the output of the simulation model was virtually identical and in current practice patients stay within their provider's team.) Once the CA and an exam room are available, the CA assesses the patient's vital signs. After this service is complete, the patient stays in the exam room to wait for the provider as shown in Illustration 6.

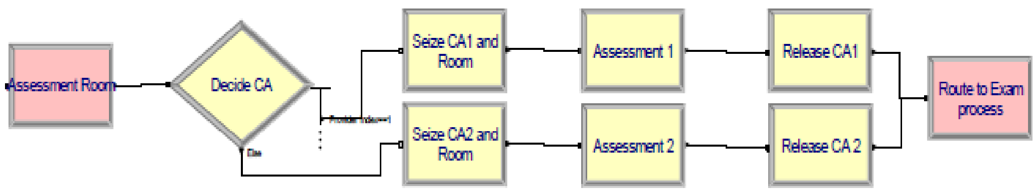


Illustration 6: Illustration of model - the service process of vital assessment

Once the provider is available, the patient receives service from the appropriate provider, determined by the patient's provider index. This process is shown in Illustration 7. The provider then examines the patient and the provider is released. Since the patient remains in the room waiting to be discharged, the patient continues to hold that resource.

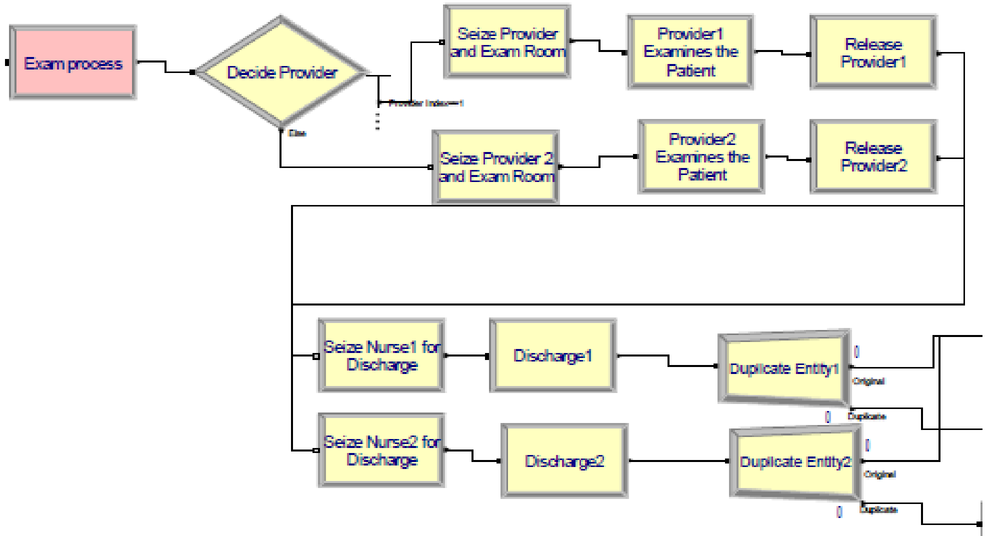


Illustration 7: Illustration of model - the service process of provider's examination

Once the provider's nurse is free, the RN discharges the patient. Even when the patient's discharge is complete, the exam room is not yet ready for the next patient. Rather, the nurse must prepare the exam room in order for it to become free. This process is shown in Illustration 8, through the use of a "duplicate entity."

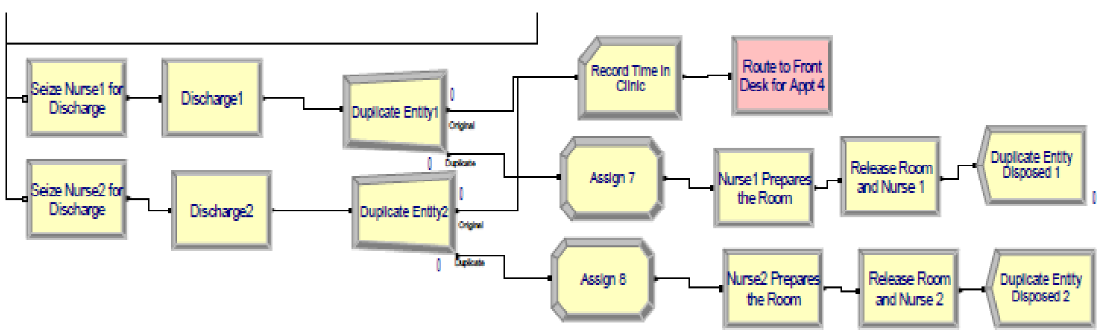


Illustration 8: Illustration of model - the service process of discharge

Once this delay for preparing the room is complete, the exam room and the nurse are free (and the artificially duplicated entity exits the system). At the same time, the real patient is routed to a CSR to possibly schedule a future appointment. The patients with a nurse appointment are handled similarly as shown in Illustration 9. They queue to be served by a nurse and similarly require an exam room. A nurse is randomly assigned to the patient. After service by the RN is complete, they are routed to the CSR for scheduling a future appointment and then exit the clinic.

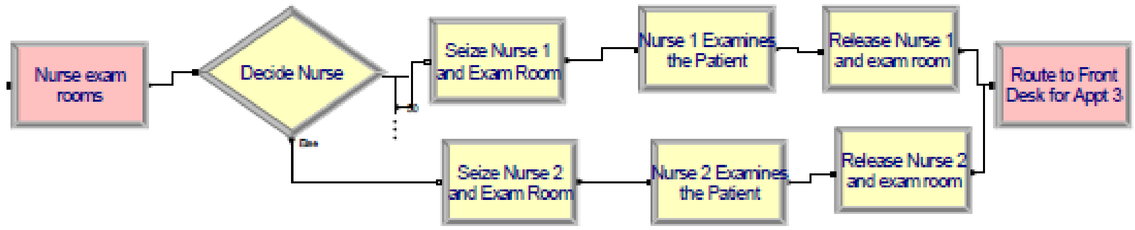


Illustration 9: Illustration of model - the service process of nurse's examination of nurse appointment patient

The last unit we model is the "distraction" unit. This is a non-negligible unit as an unoccupied staff has to handle it when a distraction occurs. This staff is then unable to handle other incoming duties until the distraction is solved. Indeed, during our communication and meetings with Seton clinical staff, we understand that they were concerned about the distractions that prevent RNs from performing their job duties. Although these "distractions" need to be handled, the staff was uncomfortable with the fact that licensed professionals (specifically RNs) were handling these distractions. For modeling purposes, we considered two distractions that we think are the most important as shown in Illustration 10. Phone calls and clerical tasks "distract" an RN. We model the distractions as entities entering the system and requiring service by an RN.

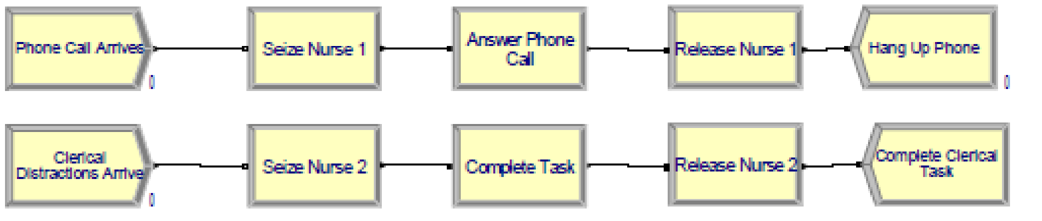


Illustration 10: Illustration of model - distractions occur in the system

We give the distractions a low priority for being served because the staff indicated that the patients in the clinic have priority over the distractions. Hence the staff will help the patient first, if a patient and a distraction come at the same time.

The entire model is shown in Illustration 11. Also a clinic diagram can be found in Illustration 12.

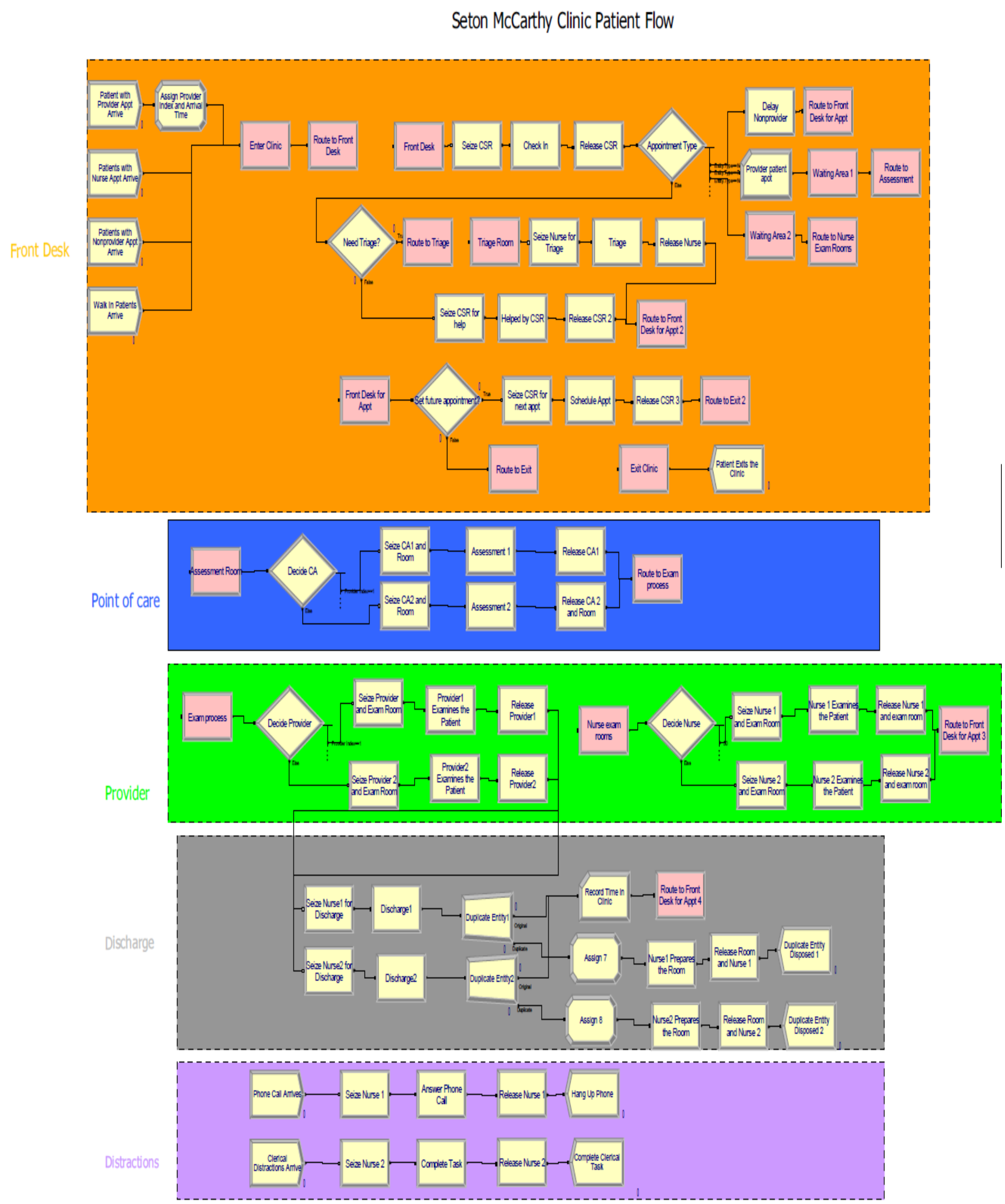


Illustration 11: Entire model

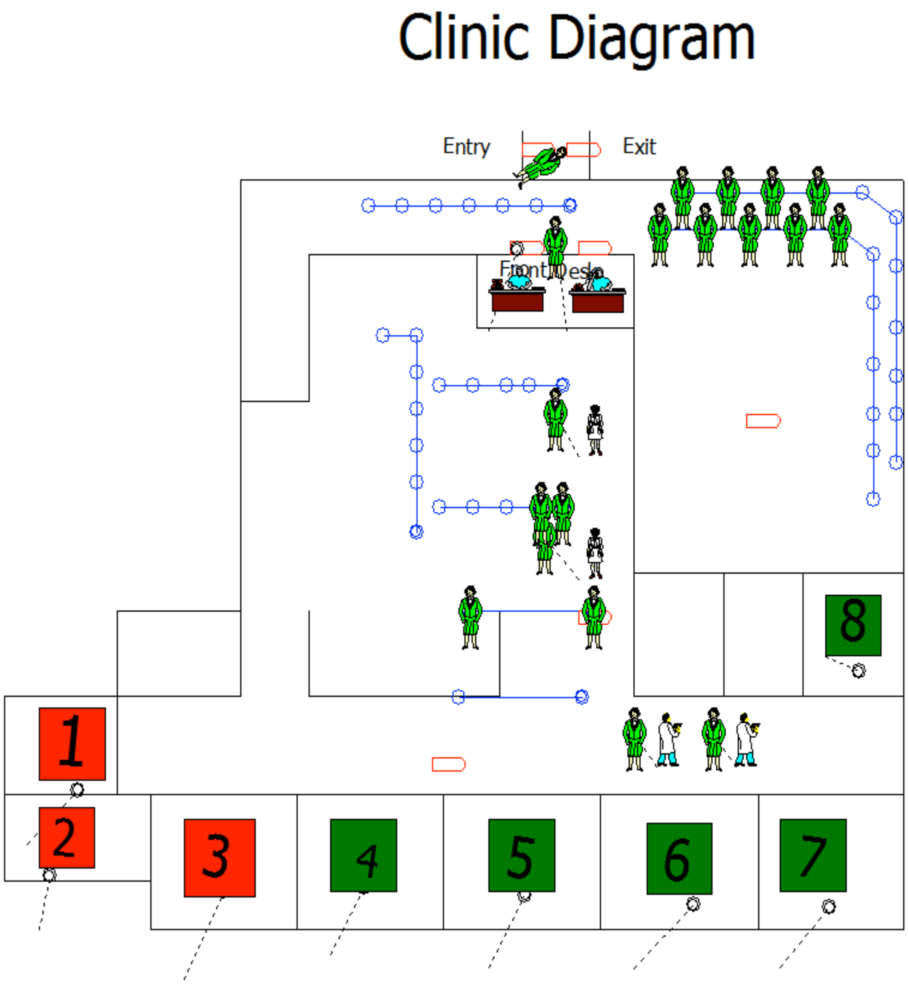


Illustration 12: Clinic diagram

3.2.2 Output of the McCarthy model

While we present the detailed simulation results for the McCarthy model in Chapter 4, we briefly highlight some results here to give the readers a flavor of the results that the simulation model yields. The results are based on 1000 replications. Each replication is a terminating simulation. The system starts empty and idle. The arrival process terminates after 240 minutes, and we run the system until the clinic is empty.

As depicted in Table 10, the 95% confidence interval (C.I.) of the mean total cycle time is 91.51±1.49 minutes. The 95% C.I. of expected waiting time for vitals is 18.16±0.84 minutes. The confidence interval for the mean waiting time for the provider's exam is 26.47±0.50 minutes and the expected waiting time for discharge has a 95% confidence interval of 10.39±0.46 minutes.

McCarthy (2-provider) Output (minutes)	Lower bound of 95% C.I.	Upper bound of 95% C.I.
Total cycle time	90.02	93.00
Waiting time for vitals	17.32	19.00
Waiting time for provider's exam	25.97	26.97
Waiting time for discharge	9.93	10.85

Table 10: 95% confidence intervals for the means of four performance measures for the McCarthy model (inter-arrival time model)

3.2.3 Model of the Kozmetsky site

For the Kozmetsky site, we create two different versions of the model: one for the days when there are two CAs, two RNs and two providers (Monday, Tuesday, Thursday and Friday afternoon, denoted MTTh) and another for the days when there are three CAs, three RNs and three providers (Wednesdays and Friday mornings, denoted WF). So the corresponding input we use for each version is shown in Table 11 and Table 12 as follows.

Input	**Distribution**
Inter-arrival time of patients (minutes)	-0.5 + GAMM(7.48, 0.95)
Process time of vital assessment (minutes)	0.5 + ERLA(1.79, 4)
Process time of provider's exam (minutes)	1.5+ERLA(3.81,3)
Process time of nurse's discharge (minutes)	-0.5 + ERLA(9.68,1)
Number of CAs	3
Number of providers	3
Number of nurses	3
Number of exam rooms	8

Table 11: Input data of 3-provider day in the Kozmetsky model (inter-arrival time model)

Input	Distribution
Inter-arrival time of patients (minutes)	-0.5 + 36 * BETA(0.538, 1.38)
Process time of vital assessment (minutes)	0.5 + ERLA(1.79, 4)
Process time of provider's exam (minutes)	1.5+ERLA(3.81,3)
Process time of nurse's discharge (minutes)	-0.5+ERLA(9.68,1)
Number of CAs	2
Number of providers	2
Number of nurses	2
Number of exam rooms	8

Table 12: Input data of 2-provider day in the Kozmetsky model (inter-arrival time model)

Patient flow at the Kozmetksy site is the same as that we describe above for the McCarthy site. Our model simulates one day at the clinic (from 8:45am to 12:45pm or 240 minutes).

3.2.4 Output of the Kozmetsky model

As shown in Table 13, for a 3-provider day at the Kozmetsky site, the 95% confidence interval of the mean total cycle time is 78.59±1.13 minutes. The 95% C.I. of expected waiting time for vitals is 25.29±0.79 minutes. The confidence interval of the mean waiting time for the provider's exam is 7.85±0.23 minutes and the expected waiting time for discharge has a 95% confidence interval of 7.94±0.25 minutes.

Kozmetsky (3-provider) Output (minutes)	Lower bound of 95% C.I.	Upper bound of 95% C.I.
Total cycle time	77.46	79.72
Waiting time for vitals	24.50	26.08
Waiting time for provider's exam	7.62	8.08
Waiting time for discharge	7.69	8.19

Table 13: 95% confidence intervals for the means of four performance measures for the 3-provider day in the Kozmetsky model (inter-arrival time model)

When there are only two providers teaming with two CAs and two RNs, the 95% C.I. of total expected cycle time shrinks to 61.21±0.87 minutes. The analogous confidence interval for the mean waiting time for assessing vital signs also shrinks to 8.10±0.46 minutes. The 95% confidence interval for the mean waiting time for the provider's exam increases to 8.75±0.33 minutes and the confidence interval for the expected waiting time for discharge also increases to 12.43±0.38 minutes as listed in Table 14.

Kozmetsky (2-provider) Output (minutes)	Lower bound of 95% C.I.	Upper bound of 95% C.I.
Total cycle time	60.34	62.08
Waiting time for vitals	7.64	8.56
Waiting time for provider's exam	8.42	9.08
Waiting time for discharge	12.05	12.81

Table 14: 95% confidence intervals for the means of four performance measures for the 2-provider day in the Kozmetsky model (inter-arrival time model)

The reason for the above change is that since the number of exam rooms is fixed (eight in total), each provider has four exam rooms on a 2-provider day while having only two or three exam rooms on a 3-provider day. When there are more exam rooms available on a 2-provider day, the patient needs to wait for less time to be brought into an exam room by a CA. The arrival process for the 2-day model and the 3-day model are such that they have the same provider-to-patient ratio. And, the rate at which the provider serves a patient is the same in both models. Thus, the patient has to wait longer in the exam room to see the provider and to be discharged.

3.2.5 Model of the Topfer site

Each week, the Topfer site has three days as a 3-provider day (Tuesday, Wednesday and Thursday) and two days (Monday and Friday) as a 2-provider day. The sets of inputs for both the 3-provider day and 2-provider day for the Topfer site are shown in Tables 15 and 16. The process of the patient flow is exactly the same as for the McCarthy site.

Input	**Distribution**
Inter-arrival time of patients (minutes)	-0.5 + GAMM(5.53, 1.26)
Process time of vital assessment (minutes)	1.5 + ERLA(2.74, 2)
Process time of provider's exam (minutes)	0.5+ERLA(4.18,3)
Process time of nurse's discharge (minutes)	-0.5+ERLA(7.62,1)
Number of CAs	3
Number of providers	3
Number of nurses	3
Number of exam rooms	9

Table 15: Input data of 3-provider day in the Topfer model (inter-arrival time model)

Input	Distribution
Inter-arrival time of patients (minutes)	-0.5 + GAMM(5.8, 1.39)
Process time of vital assessment (minutes)	1.5 + ERLA(2.74, 2)
Process time of provider's exam (minutes)	0.5+ERLA(4.18,3)
Process time of nurse's discharge (minutes)	-0.5+ERLA(7.62,1)
Number of CAs	2
Number of providers	2
Number of nurses	2
Number of exam rooms	9

Table 16: Input data of 2-provider day in the Topfer model (inter-arrival time model)

3.2.6 Output of the Topfer model

As shown in Table 17, the 95% confidence interval for mean total cycle time on a 3- provider day is 54.37±0.67 minutes. The 95% C.I. for mean waiting time for vitals is 9.63±0.40 minutes. The confidence interval for expected waiting time for provider's exam is 7.10±0.19 minutes and the 95% confidence interval for the expected waiting time for discharge is 5.89± 0.18 minutes.

Topfer (3-provider) Output (minutes)	Lower bound of 95% C.I.	Upper bound of 95% C.I.
Total cycle time	53.70	55.04
Waiting time for vitals	9.23	10.03
Waiting time for provider's exam	6.91	7.29
Waiting time for discharge	5.71	6.07

Table 17: 95% confidence intervals for the means of four performance measures for 3-provider day in the Topfer model (inter-arrival time model)

When there are only two providers teaming with two CAs and two RNs, the 95% C.I. for the mean total cycle time increases to 59.54±0.88 minutes. The analogous C.I. for mean waiting time for assessment of vital signs shrinks to 7.69±0.40 minutes. The 95% C.I. for the expected waiting time for the provider's exam increases to 11.95±0.33 minutes and the C.I. for the mean waiting time for discharge also increases to 9.94±0.32 minutes as listed in Table 18.

Topfer (2-provider) Output (minutes)	Lower bound of 95% C.I.	Upper bound of 95% C.I.
Total cycle time	58.66	60.42
Waiting time for vitals	7.29	8.09
Waiting time for provider's exam	11.62	12.28
Waiting time for discharge	9.62	10.26

Table 18: 95% confidence intervals for the means of four performance measures for 2-provider day in the Topfer model (inter-arrival time model)

The reason for the above difference is that since the number of exam rooms (nine in total) is fixed, each provider will have four and half exam rooms on a 2-provider day and have three exam rooms on a 3-provider day. When there are more exam rooms available on a 2-provider day, the patient needs to wait for less time to be brought into an exam room by a CA and then must wait longer in the exam room to see the provider and to be discharged. The total increment in the expected waiting time for the provider's exam and for discharge exceeds the time decrease in waiting for vital assessment. Hence, there is a bit of an increase in the mean total cycle time in a 2-provider day.

3.2.7 Effectiveness of the model

To determine whether our modeling approach is effective in reflecting the real system, we compare the model output to the performance of the actual system. Since the most important measurement is the total cycle time, we use total cycle time as the indicator to see if the output of the model is similar to the real total cycle time from data collected onsite. Figure 6 graphically shows the total cycle time in the clinic, i.e., the difference between when the patient leaves the clinic and the patient's arrival time. As the figure shows, only the confidence intervals created from the model output of the Kozmetsky 2- provider day contain the average value from the actual system. Next we use an alternative method to model the arrival process and see if we can improve the effectiveness of our model in reproducing observed cycle times.

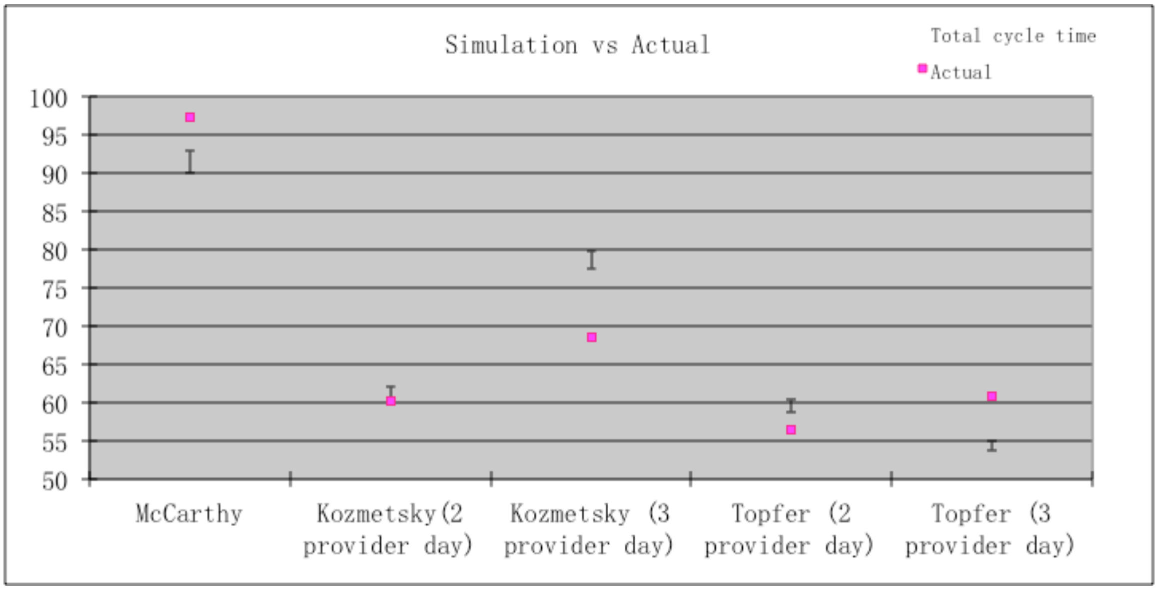


Figure 6: Comparison between simulation result and actual data (inter-arrival time). The y-axis displays cycle time in minutes.

3.3 MODEL WITH DEVIATION OF ARRIVAL FROM APPOINTMENT TIME

In this section, we model the arrival process using the distribution governing the deviation of the arrival time from the patient's appointment time. As we describe in Chapter 2, for each clinical site, we separate the deviations into two categories: being early and being late. The probability for a patient to be early (late) is calculated for each clinical site. At the McCarthy site, 4.01% of patients are punctual, 70.85% of patients are early and 25.14% of patients are late. At the Kozmetsky site, 6.41% of patients are punctual, 71.26% of patients are early and 22.33% of patients are late for their appointments. At the Topfer site, 7.31% of patients are punctual, 72.87% of patients are early and 19.81% of patients are late.

Appointments are scheduled at fixed intervals of 15 minutes throughout the day, and the actual arrival time of each patient is modeled as the appointment time delayed by the amount of the deviation (a negative value stands for early deviation, and a positive value stands for late deviation). Thus, when a patient enters the system, the model decides whether the patient is on time or is early for his/her appointment or is late for his/her appointment according to the probability we estimated for this clinical site. If the patient is early, then the distribution of deviation of early arrivals is used to model the arrival time of the patient. If the patient is assigned to be late, then the distribution of deviation of late arrivals is used to model the arrival time of the patient. Since we cannot model early deviation as a delay of a negative amount of time, we use the following method: we offset the schedule time by a certain amount of time, and then introduce a corresponding delay to cancel off the offset. For example, the largest early deviation is 120 minutes for the McCarthy site. Hence we offset all schedules in McCarthy two hours early. At the McCarthy site, the probability of the patient being punctual is 0.0401. If the patient arrives on time, then in our simulation model the patient is delayed for 120 minutes before released to enter the system. The probability of a patient being early is 0.7085. If a patient arrives early, the time that the system delays this patient before releasing him/her to the system is a random variable with the distribution of 119.001-EXPO(18.1) (less than 120 minutes). The probability of a patient being late is 0.2514. If a patient arrives late, the time that the system delays the patient is a random variable with a distribution of 120.5+GAMMA(6.99,1.28) (more than 120 minutes). It is the same for the other two sites except that the baseline of time delayed is 180 minutes for the Kozmetsky site and 240 minutes for the Topfer site because of the largest early deviations in these two sites. The ultimate effect is that patients have appointments at 15 minute intervals throughout the day and we model deviations from those times. The model is shown in Illustration 13.

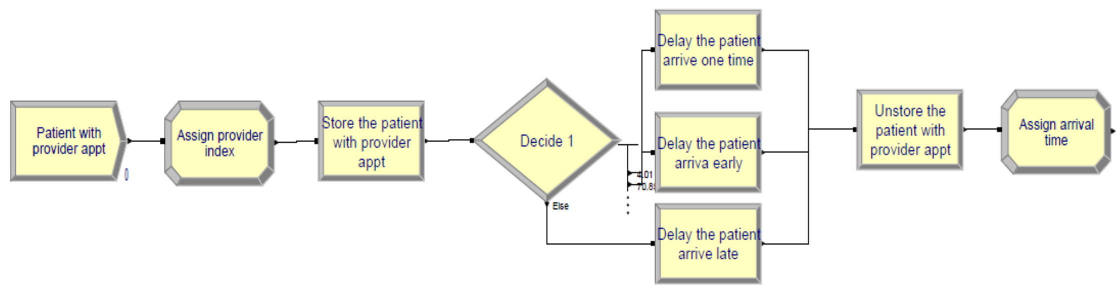


Illustration 13: Illustration of model - modeling arrival process using deviation from appointment time

3.3.1 Model of the McCarthy site

The input data in our alternative model for the McCarthy site is shown in Table 19. All the distributions used in the model are the same as used previously (shown in Table 9) except that the distribution of arrivals are modeled differently.

Input	Distribution
Deviation of patient (Early) (minutes)	119.001-EXPO(18.1)
Deviation of patient (Late) (minutes)	120.5+GAMM(6.99,1.28)
Process time of vital assessment (minutes)	1.5+ERLA(1.66,4)
Process time of provider's exam (minutes)	5+ERLA(5.62,3)
Process time of nurse's discharge (minutes)	-0.001+ERLA(10.8,1)
Number of CAs	2
Number of providers	2
Number of nurses	2
Number of exam rooms	8

Table 19: Input data of the McCarthy model (deviation of arrival from appointment time)

3.3.2 Output of the McCarthy model

McCarthy (2-provider) Output (minutes)	Lower bound of 95% C.I.	Upper bound of 95% C.I.
Total cycle time	95.04	98.14
Waiting time for vitals	22.19	23.85
Waiting time for provider's exam	26.11	26.93
Waiting time for discharge	5.31	5.83

Table 20: 95% confidence intervals for the means of four performance measures for the McCarthy model (deviation of arrival from appointment time)

As depicted in Table 20, the 95% confidence interval of expected total cycle time is 96.59±1.55 minutes. The 95% C.I. for mean waiting time for vitals is 23.02±0.83 minutes. The confidence interval for the mean time for waiting for provider’s exam is 26.52±0.41 minutes and the waiting time for discharges ranges within the interval of 5.57±0.26 minutes.

3.3.3 Model of the Kozmetsky site

The input data in our alternative model for the Kozmetsky site for a 3-provider day and a 2-provider day are depicted in Table 21 and Table 22, respectively. All the distributions used in the model are the same as used previously (shown in Table 11 and Table 12) except that the distribution of arrivals is modeled differently using deviation of patient’s arrival time from appointment time.

Input	Distribution
Deviation of patient (Early) (minute)	179.001-EXPO(20.9)
Deviation of patient (Late) (minute)	180.5 + GAMM(8.12, 1.19)
Process time of vital assessment (minute)	0.5 + ERLA(1.79, 4)
Process time of provider's exam (minute)	1.5+ERLA(3.81,3)
Process time of nurse's discharge (minute)	-0.5+ERLA(9.68,1)
Number of CAs	3
Number of providers	3
Number of nurses	3
Number of exam rooms	8

Table 21: Input data of 3-provider day in the Kozmetsky model (deviation of arrival from appointment time)

Input	Distribution
Deviation of patient (Early) (minute)	179.001-EXPO(20.9)
Deviation of patient (Late) (minute)	180.5 + GAMM(8.12, 1.19)
Process time of vital assessment (minute)	0.5 + ERLA(1.79, 4)
Process time of provider's exam (minute)	1.5+ERLA(3.81,3)
Process time of nurse's discharge	-0.5+ERLA(9.68,1)
Number of CAs	2
Number of providers	2
Number of nurses	2
Number of exam rooms	8

Table 22: Input data of 2-provider day in the Kozmetsky model (deviation of arrival from appointment time)

3.3.4 Output of the Kozmetsky model

As shown in Table 23, for a 3-provider day, the 95% C.I. for mean total cycle time is 69.34±0.83 minutes. The 95% C.I. for mean waiting time for vitals is 26.09±0.57

minutes. The analogous confidence interval for the mean waiting time for provider's exam is 4.48±0.19 minutes and the 95% confidence interval for the expected waiting time for discharge is 4.68±0.12 minutes.

Kozmetsky (3-provider) Output (minutes)	Lower bound of 95% C.I.	Upper bound of 95% C.I.
Total cycle time	68.51	70.17
Waiting time for vitals	25.52	26.66
Waiting time for provider's exam	4.29	4.67
Waiting time for discharge	4.56	4.80

Table 23: 95% confidence intervals for the means of four performance measures for the 3-provider day in the Kozmetsky model (deviation of arrival from appointment time)

As shown in Table 24, when there are only two providers teaming with two CAs and two RNs, the 95% C.I. for mean total cycle time shrinks to 55.39±0.73 minutes. The 95% C.I. for mean waiting time for vital assessment also shrinks to 9.02±0.37 minutes. The analogous confidence interval for the mean waiting time for provider's exam increases to 7.99±0.20 minutes, and the 95% confidence interval for the expected waiting time for discharge also increases to 7.12±0.22 minutes. Note that the trend of changes follows the same pattern as shown in Table 13 and 14.

Kozmetsky (2-provider) Output (minutes)	Lower bound of 95% C.I.	Upper bound of 95% C.I.
Total cycle time	54.66	56.12
Waiting time for vitals	8.65	9.39
Waiting time for provider's exam	7.79	8.19
Waiting time for discharge	6.90	7.34

Table 24: 95% confidence intervals for the means of four performance measures for the 2-provider day in Kozmetsky model (deviation of arrival from appointment time)

3.3.5 Model of the Topfer site

The input data for a 3-provider day and a 2-provider day at the Topfer site are shown in Table 25 and Table 26, respectively. All the distributions used in the model are the same as used previously (shown in Table 15 and Table 16) except that the distribution of arrivals is modeled differently using deviation of a patient’s arrival time from appointment time.

Input	**Distribution**
Deviation of patient (Early) (minute)	239.001-EXPO(18.5)
Deviation of patient (Late) (minute)	240.5+GAMM(7.37,1.36)
Process time of vital assessment (minute)	1.5 + ERLA(2.74, 2)
Process time of provider's exam (minute)	0.5+ERLA(4.18,3)
Process time of nurse's discharge (minute)	-0.5+ERLA(7.62,1)
Number of CAs	3
Number of providers	3
Number of nurses	3
Number of exam rooms	9

Table 25: Input data of 3-provider day in the Topfer model (deviation of arrival from appointment time)

Input	**Distribution**
Deviation of patient (Early) (minute)	239.001-EXPO(18.5)
Deviation of patient (Late) (minute)	240.5+GAMM(7.37,1.36)
Process time of vital assessment (minute)	1.5 + ERLA(2.74, 2)
Process time of provider's exam (minute)	0.5+ERLA(4.18,3)
Process time of nurse's discharge (minute)	-0.5+ERLA(7.62,1)
Number of CAs	2
Number of providers	2
Number of nurses	2
Number of exam rooms	9

Table 26: Input data of 2-provider day in the Topfer model (deviation of arrival from appointment time)

3.3.6 Output of the Topfer model

As shown in Table 27, for a 3-provider day, the 95% confidence interval for mean total cycle time is 61.89±0.68 minutes. The 95% C.I. for mean waiting time for vitals is 15.58±0.42 minutes. The analogous confidence interval for the mean waiting time for provider's exam is 8.27±0.13 minutes and the 95% confidence interval for the expected waiting time for discharge is 3.00±0.18 minutes.

Topfer (3-provider) Output (minutes)	Lower bound of 95% C.I.	Upper bound of 95% C.I.
Total cycle time	61.21	62.57
Waiting time for vitals	15.16	16.00
Waiting time for provider's exam	8.14	8.40
Waiting time for discharge	2.93	3.09

Table 27: 95% confidence intervals for the means of four performance measures for the 3-provider day in the Topfer model (deviation of arrival from appointment time)

When there are only two providers teaming with two CAs and two RNs, the 95% C.I. for mean total cycle time shrinks to 53.41±0.63 minutes. The 95% C.I. for mean waiting time for vital assessment also shrinks to 4.33±0.22 minutes. The analogous confidence interval for the mean waiting time for provider's exam increases to 11.58±0.29 minutes and the 95% confidence interval for the expected waiting time for discharge also increases to 4.06±0.14 minutes. The trend of changes follows the same pattern as shown in Table 17 and 18 but the amount of the decrease in the waiting time to have a patient's vitals assessed is now greater than the increment in the time spent waiting for the provider's exam and for the discharge, which leads to a decrease in mean total cycle time in a 2-provider day.

Topfer (2-provider) Output (minutes)	Lower bound of 95% C.I.	Upper bound of 95% C.I.
Total cycle time	52.78	54.04
Waiting time for vitals	4.11	4.55
Waiting time for provider's exam	11.29	11.87
Waiting time for discharge	3.92	4.20

Table 28: 95% confidence intervals for the means of four performance measures for the 2-provider day in the Topfer model (deviation of arrival from appointment time)

3.3.7 Effectiveness of the model

In this section, we compare the model output to the performance of the actual system, as shown in Figure 7. We observe that the confidence intervals obtained from the model output of all sites contains the average value of the actual system except the model for a 3- provider day at the Kozmetsky site.

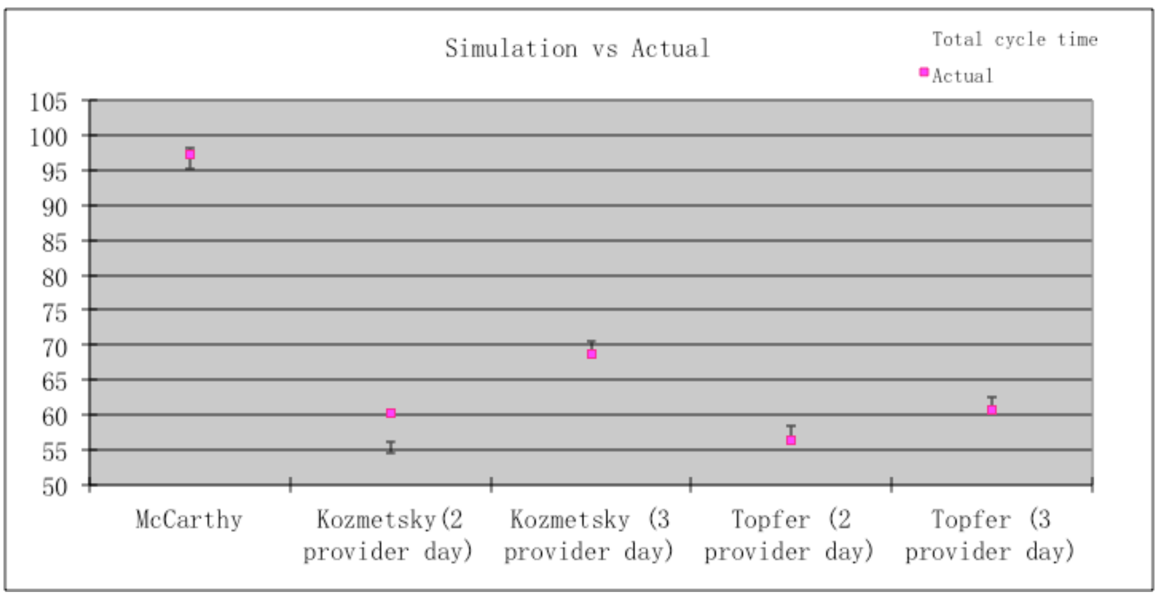


Figure 7: Comparison between simulation result and actual data (deviation of arrival from appointment time). The y-axis displays cycle time in minutes.

As we can see by comparing Figure 6 and 7, using deviation of arrivals from appointment time to model the patients' arrival process better reflects the real system. This indeed reveals the fact that the time that a patient arrives critically depends on his appointment time. Hence, based on the probability that the patient arrives early or arrives late, the arrivals of the patients in the model reflects the real system better than just using inter-arrival times without considering the patients' appointment times.

Chapter 4 Alternative system designs

In this chapter, we conduct sensitivity analysis to assess the performance of the system at each of the three clinics under two alternative system designs. We use the models that use deviation of arrival time from appointment to model the patient arrival process. The models we analyzed in Chapter 3 serve as our base scenario for each clinic. We consider two other system designs that involve "sharing an additional CA" and "sharing an additional RN." These two alternative system designs are the most realistic modifications to the system's resources. The clinical sites are fixed and remodeling the building, e.g., to add another exam room is not realistic. And, adding a CA or an RN is more realistic than adding another provider given the overall system design.

4.1 ALTERNATIVE SYSTEM DESIGNS FOR THE MCCARTHY SITE

We begin with the McCarthy site, leaving all other system parameters fixed as we describe in Chapter 2 and 3, and we add a single CA, or a single RN, that the three providers share. Figures 8-11 show the results for the McCarthy simulation model, again using 1000 replications for the four performance measures we consider: the expected values of the total cycle time, the waiting time for having vital signs assessed, the waiting time for the provider, and the waiting time for discharge. As shown in Figure 8, adding a CA does not significantly change the total cycle time but adding a nurse decreases the total cycle time by about 20 minutes. In particular, the 95% C.I. of the mean total cycle time decreases from 96.59±1.03 minutes to 76.35±1.11 minutes.

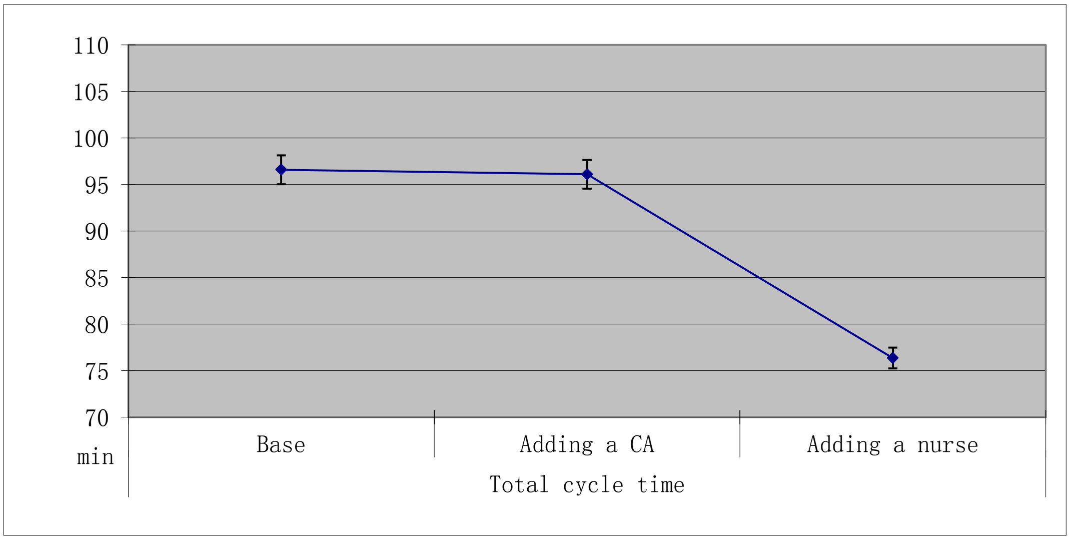


Figure 8: Sensitivity analysis of the McCarthy site (expected total cycle time)

Also, sharing an additional nurse causes a decrease in expected waiting time for assessing vital signs of about 9 minutes as shown in Figure 9. The 95% C.I. for expected waiting time for assessing vital signs decreases from 23.02±0.83 minutes to 13.62±0.54 minutes. When sharing an additional nurse, the previous patient is discharged more quickly and hence the exam room becomes free for taking vital signs more quickly, hence the decrease in expected waiting time for vitals assessment.

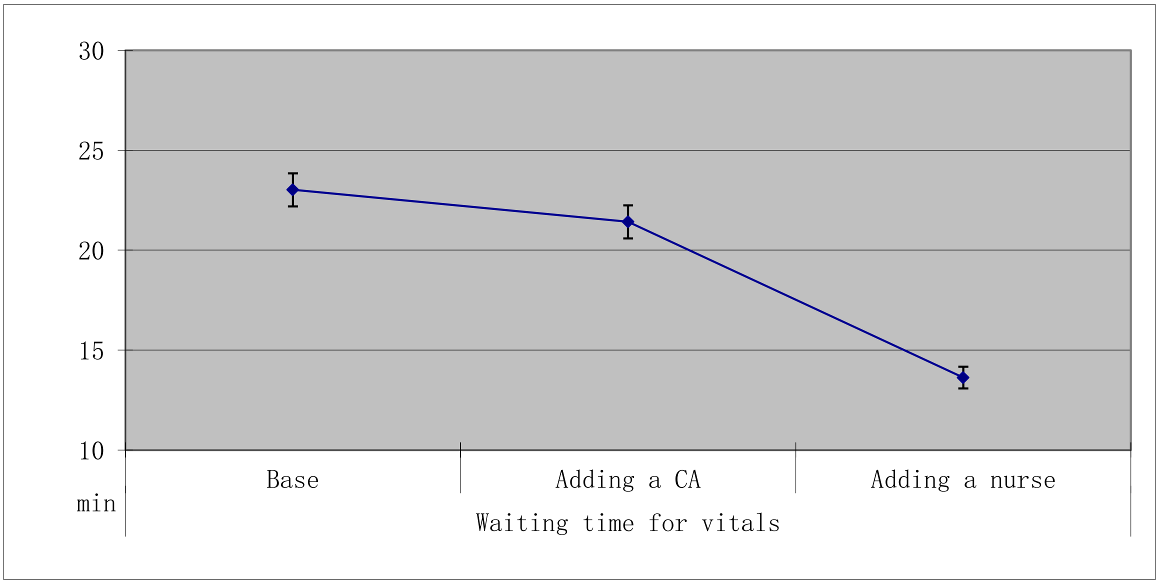


Figure 9: Sensitivity analysis of the McCarthy site (expected waiting time for vital assessment)

As shown in Figure 10, neither sharing an additional nurse nor sharing an additional CA improves the expected waiting time for the provider's exam. On the contrary, patients have to wait longer to see the provider. This is because the provider's service rate is fixed and even when the patient is brought to the exam room earlier by a CA, the patient still needs to wait for the provider to finish examining the current patient.

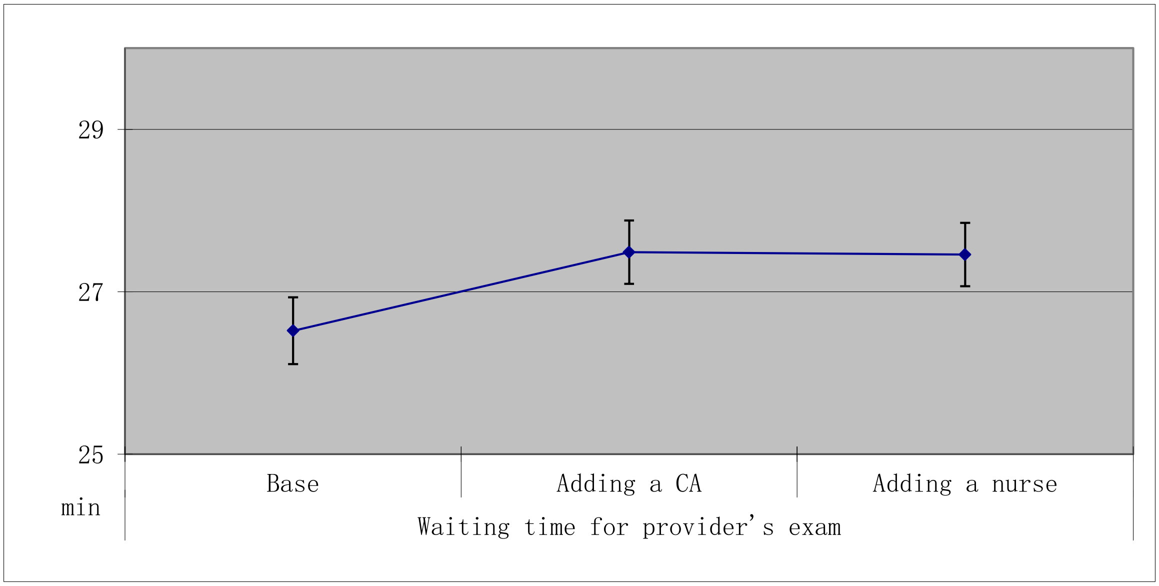


Figure 10: Sensitivity analysis of the McCarthy site (expected waiting time for provider's exam)

Sharing an additional nurse decreases the waiting time in discharge by about five minutes as shown in Figure 11. The 95% C.I. on the mean waiting time for discharge decreases from 5.57±0.26 minutes to 0.54±0.05 minutes.

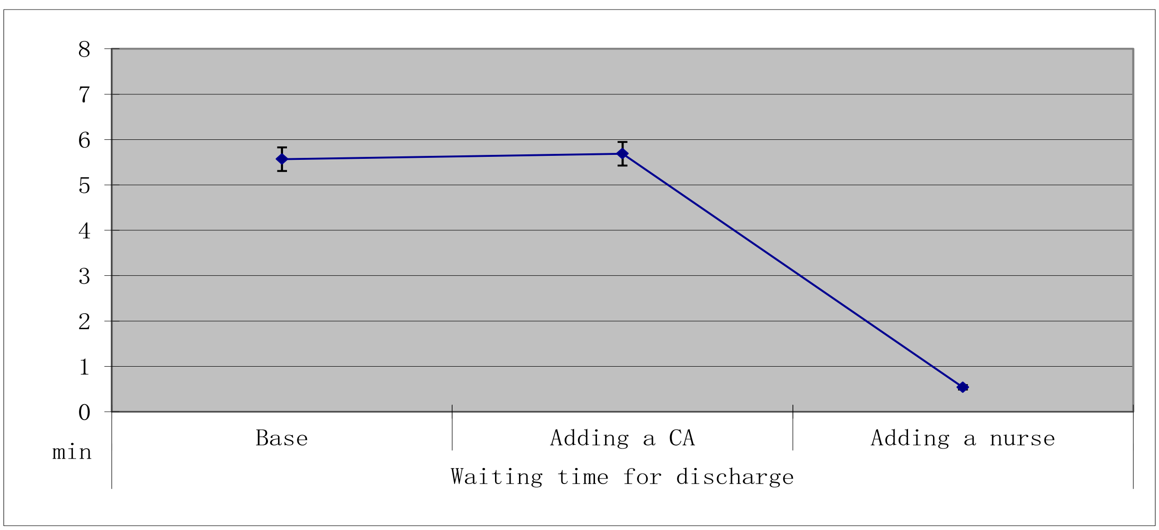


Figure 11: Sensitivity analysis of the McCarthy site (expected waiting time for discharge)

4.2 ALTERNATIVE SYSTEM DESIGNS FOR THE KOZMETSKY SITE

Figure 12 shows that on a 2-provider day at the Kozmetsky site, the effect of adding a CA and adding a nurse is similar to that for the McCarthy site. In particular, the expected total cycle time decreases about 10 minutes with the 95% C.I. decreasing from 55.39±0.73 minutes to 45.43±0.49 minutes when an additional nurse is shared.

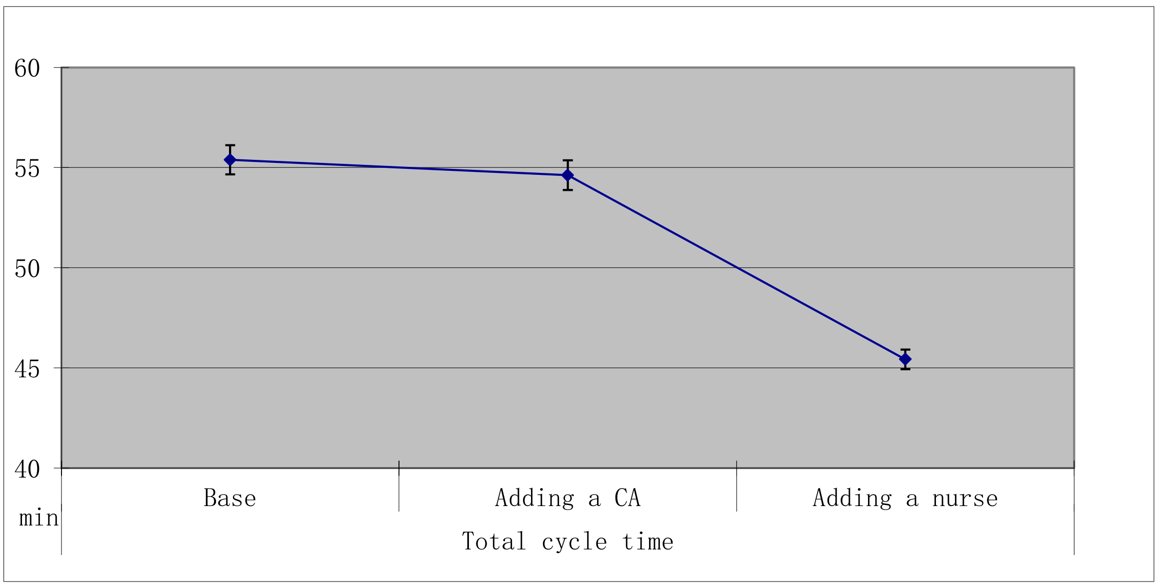


Figure 12: Sensitivity analysis of 2-provider Kozmetsky site (expected total cycle time)

Sharing an additional CA or RN decreases the expected waiting time for assessing vital signs as shown in Figure 13. The 95% C.I. for the expected time for assessing vital signs decreases from 9.02±0.37 minutes to 5.95±0.24 minutes when an additional nurse is shared. Again neither sharing an additional CA nor sharing an additional nurse improves the expected waiting time for the provider's exam. This is true for the same reason we describe in the above discussion of the McCarthy site.

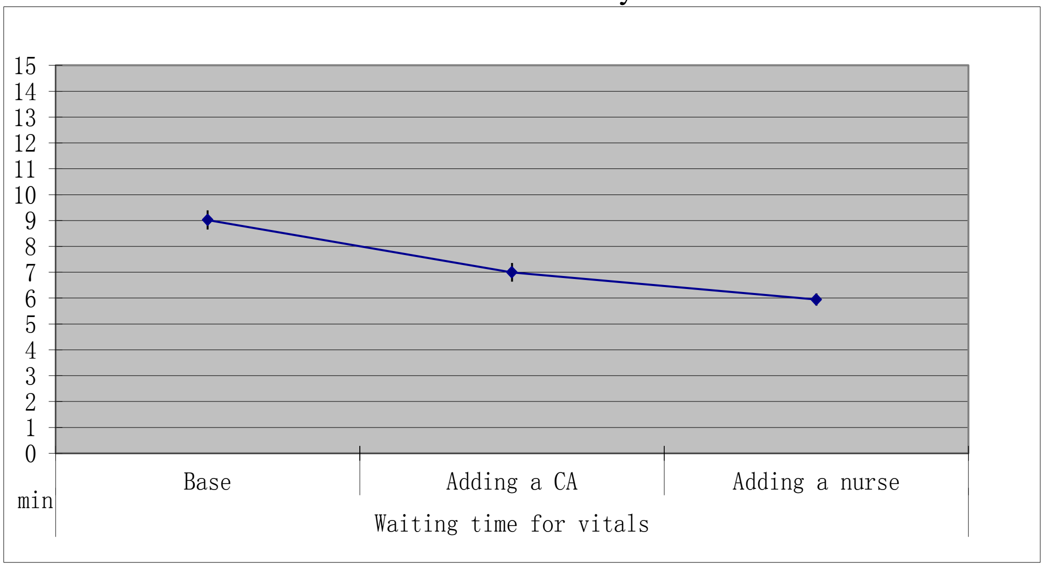


Figure 13: Sensitivity analysis of 2-provider Kozmetsky site (expected waiting time for vital assessment)

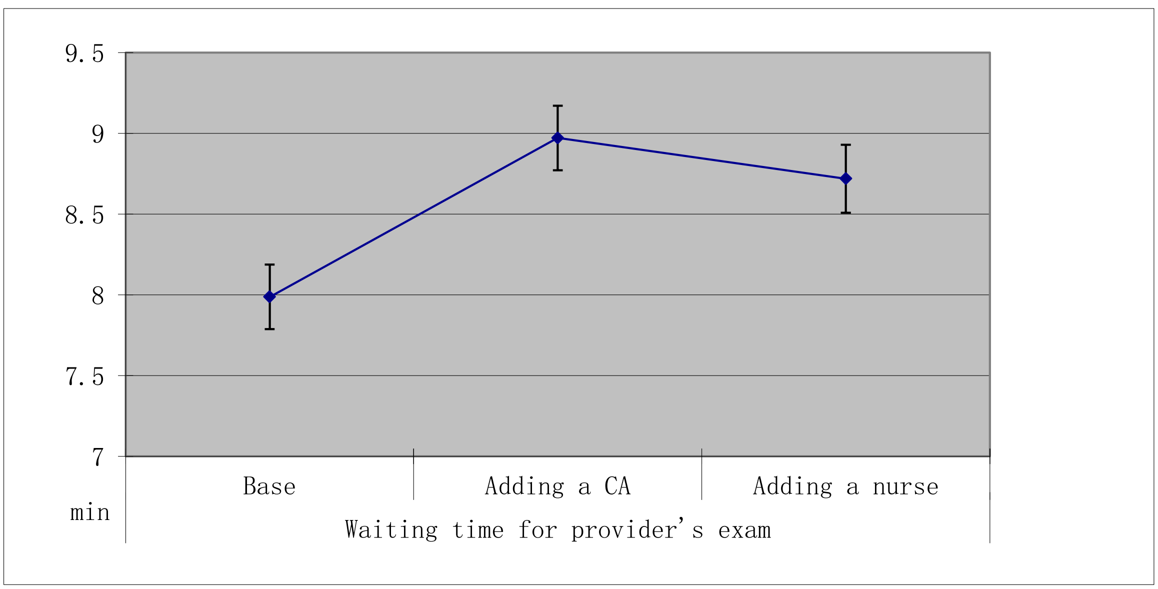


Figure 14: Sensitivity analysis of 2-provider Kozmetsky site (expected waiting time for provider's exam)

On a 2-provider day at the Kozmetsky site, sharing an additional nurse again significantly decreases the expected waiting time for discharge, as shown in Figure 15.

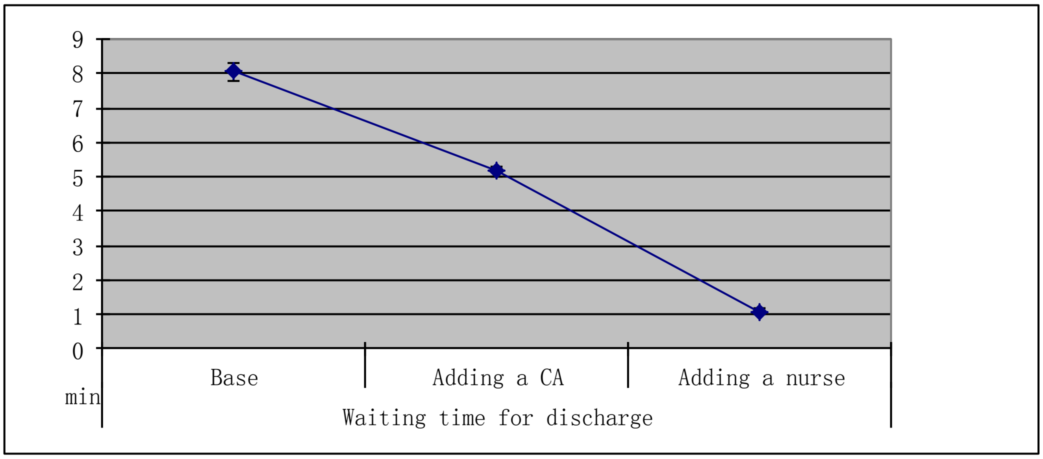


Figure 15: Sensitivity analysis of 2-provider Kozmetsky site (expected waiting time for discharge)

In a 3-provider day at the Kozmetsky site, sharing an additional nurse again leads to a decrease in expected total cycle time of about 10 minutes as shown in Figure 16. The 95% C.I. moves from 69.34±0.83 minutes to 59.05±0.69 minutes. This is because when each provider only has two or three exam rooms, the exam rooms become a scarce resource and sharing an additional nurse decreases the time for discharge and hence less time is required for a room to be ready for the next patient. Sharing an additional nurse outperforms sharing an additional CA in shortening the expected total cycle time, as shown in Figure 16.

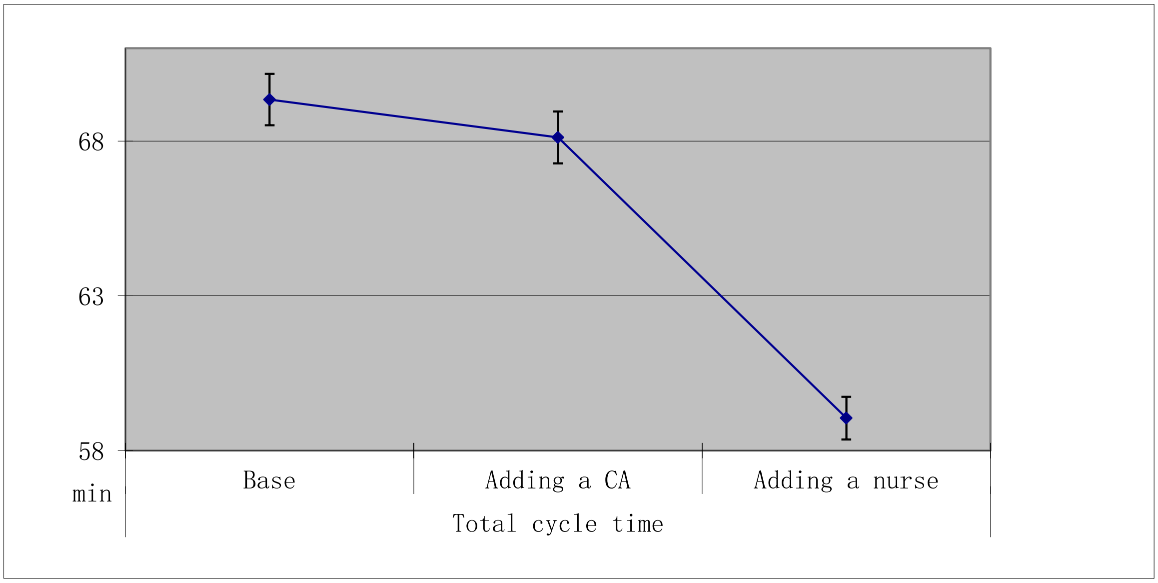


Figure 16: Sensitivity analysis of 3-provider Kozmetsky site (expected total cycle time)

Sharing an additional nurse also decreases the expected waiting time for assessing vital signs by about six minutes. This is because when the room is ready for the next patient faster, the patient has to wait less to be called in by the CAs.

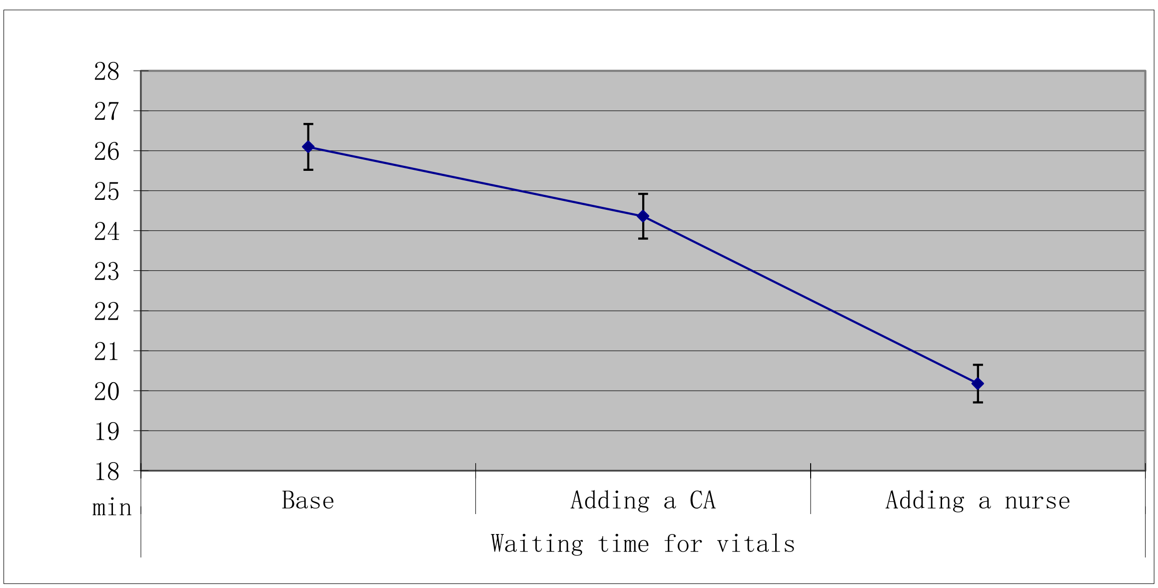


Figure 17: Sensitivity analysis of 3-provider Kozmetsky site (expected waiting time for vital assesment)

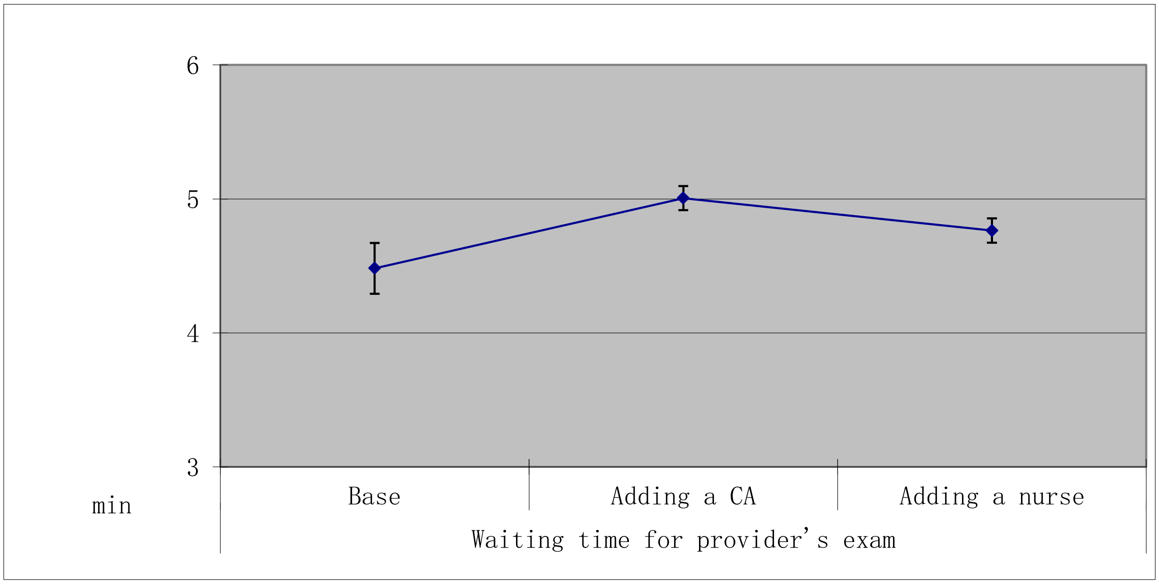


Figure 18: Sensitivity analysis of 3-provider Kozmetsky site (expected waiting time for provider's exam)

As shown in Figure 18, adding a nurse or CA does not improve the waiting time for the provider's exam. Sharing an additional nurse does improve the waiting time for discharge (three and half minutes), as shown in Figure 19. We notice that the 95% C.I. of expected waiting time is relatively small.

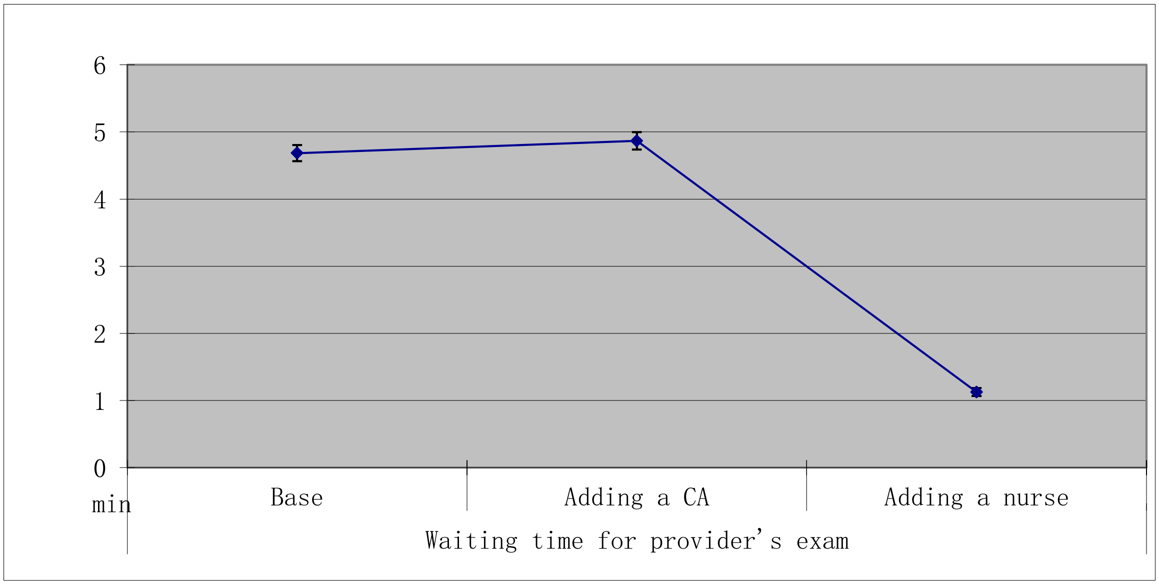


Figure 19: Sensitivity analysis of 3-provider Kozmetsky site (expected waiting time for discharge)

4.3 Alternative system designs for the Topfer site

On a 2-provider day at the Topfer, sharing an additional nurse decreases the expected total cycle time by about four minutes. The 95% C.I. moves from 53.41±0.63 minutes to 48.76±0.53 minutes as shown in Figure 20.

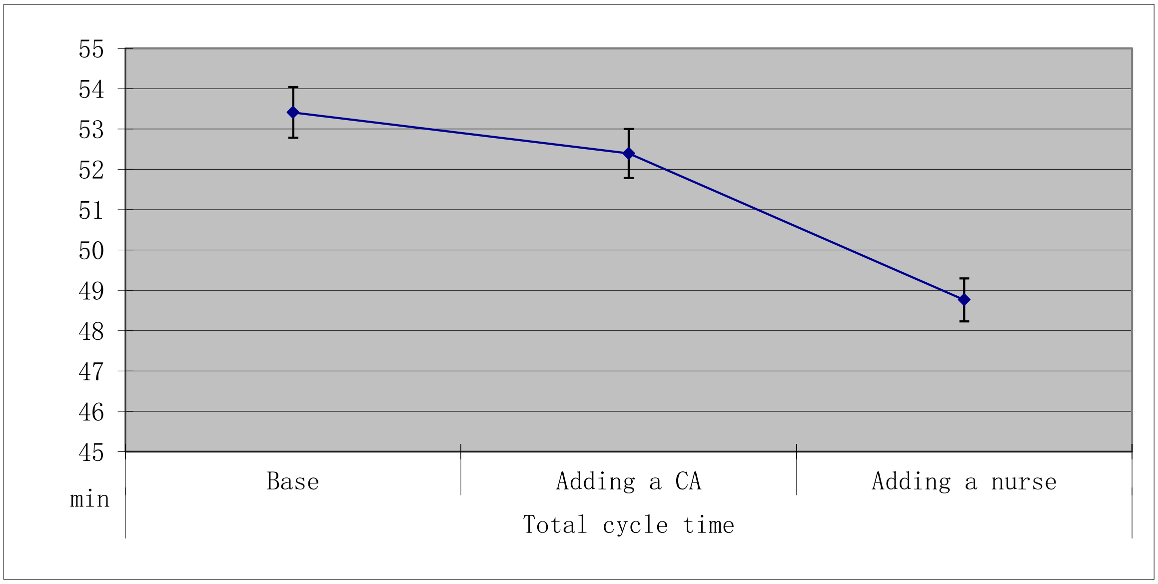


Figure 20: Sensitivity analysis of 2-provider Topfer site (expected total cycle time)

Sharing an additional nurse does not help in decreasing the expected time for assessing vital signs. In fact, adding a CA leads to the most decrease in waiting time for vital assessment. The reason is that there are nine exam rooms at the Topfer site. When there are only two providers, sharing an additional nurse to help to get the exam room ready faster will not help to decrease the waiting time for vital assessment since the patient still has to wait for an available CA to bring them in. That is why sharing an additional CA helps the most in decrease waiting time for vital assessment, as shown in Figure 21.

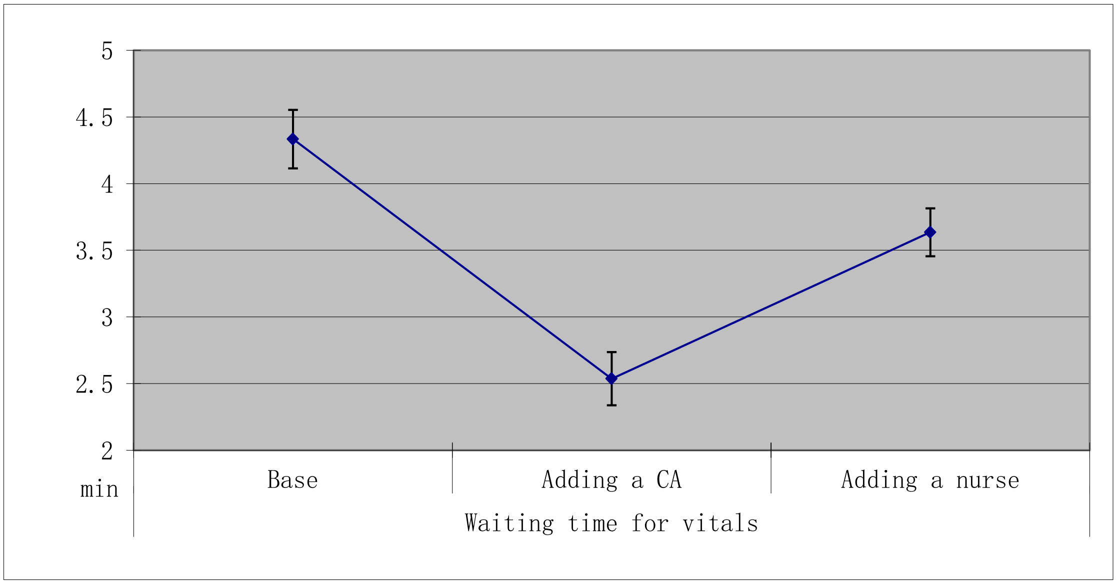


Figure 21: Sensitivity analysis of 2-provider Topfer site (expected waiting time for vital assessment)

As shown in Figure 22, neither sharing an additional CA nor sharing an additional nurse helps improve the expected waiting time for the provider's exam while it does help in decreasing the waiting time for discharge, as shown in Figure 23. We notice that the C.I. for waiting time for discharge is relatively small.

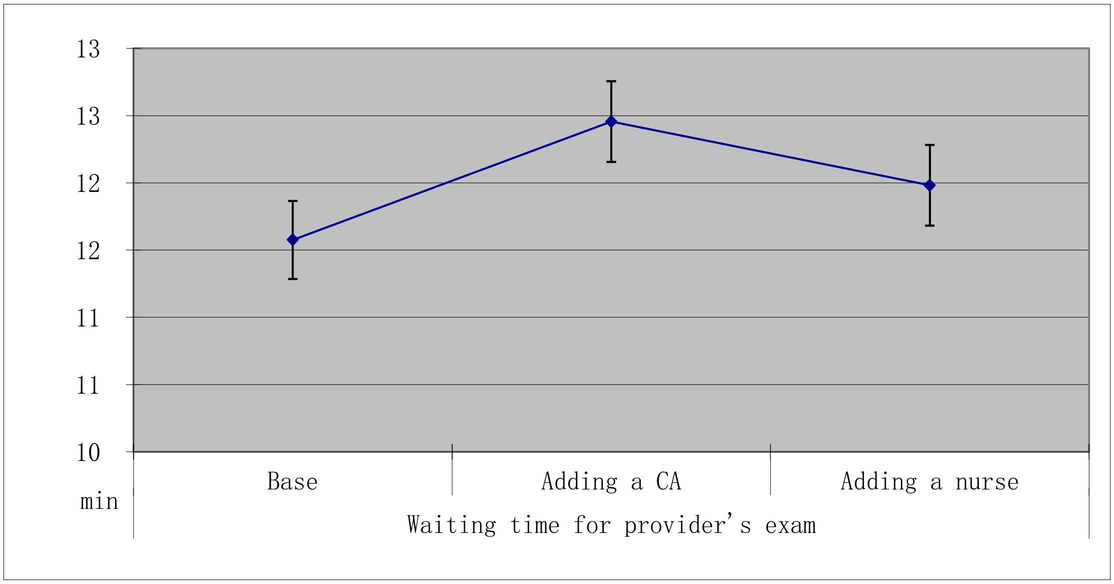


Figure 22: Sensitivity analysis of 2-provider Topfer site (expected waiting time for provider's exam)

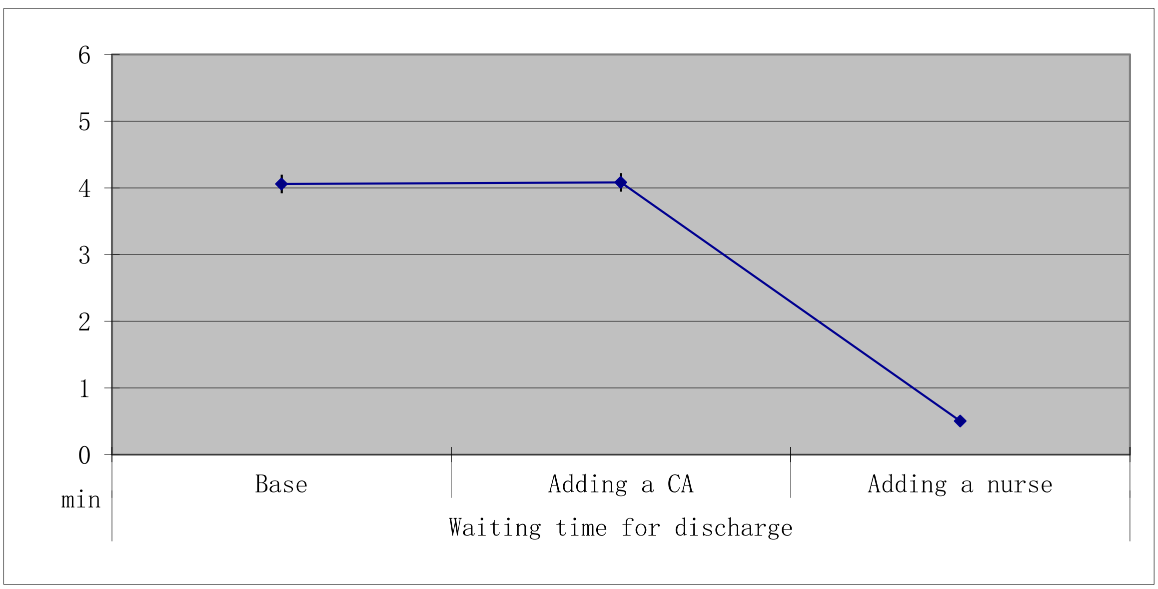


Figure 23: Sensitivity analysis of 2-provider Topfer site (expected waiting time for discharge)

In a 3-provider day at the Topfer site, adding a nurse decreases the total cycle by about five minutes and also decreases the expected waiting time for vital assessment by two minutes and the expected waiting time for discharge by about two minutes. The 95% C.I. of expected total cycle time moves from 61.89±0.68 minutes to 56.52±0.60 minutes, as shown in Figure 24. The 95% C.I. for expected waiting time for vitals moves from 15.58±0.42 minutes to 13.10±0.38 minutes, as shown in Figure 25.

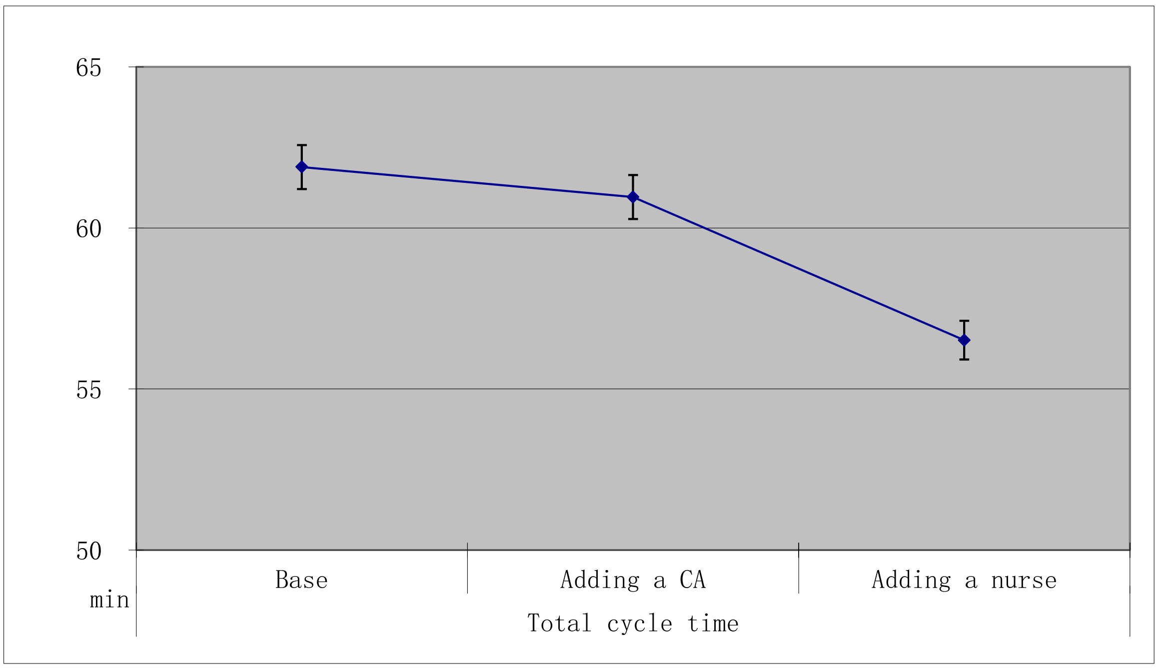


Figure 24: Sensitivity analysis of 3-provider Topfer site (expected total cycle time)

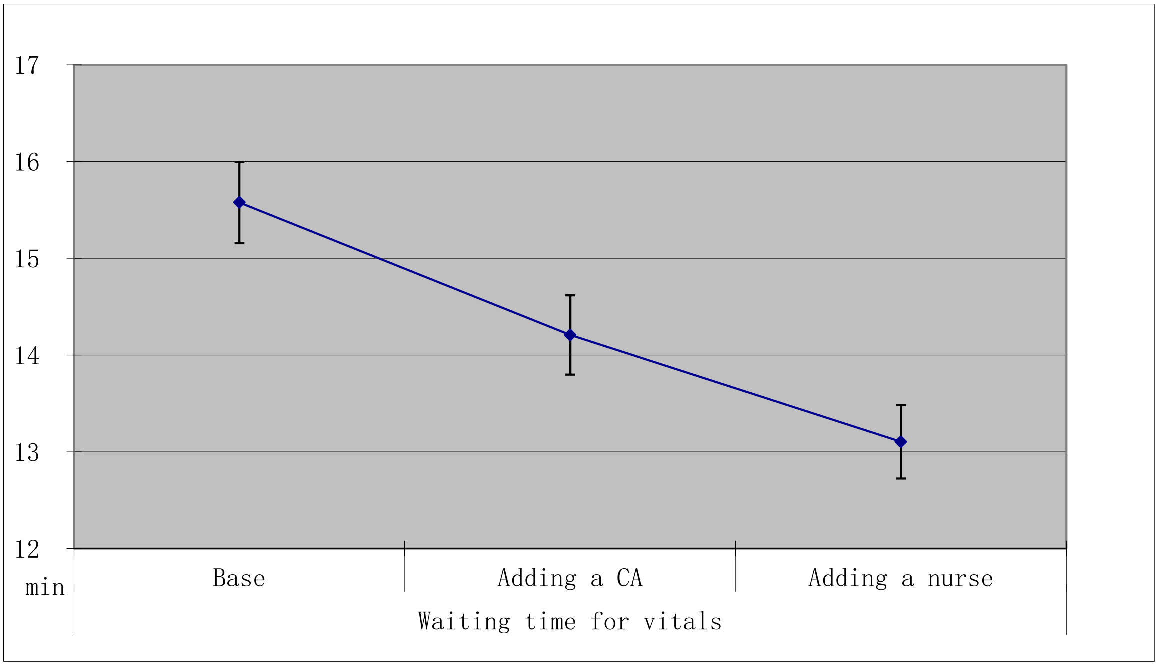


Figure 25: Sensitivity analysis of 3-provider Topfer site (expected waiting time for vital assessment)

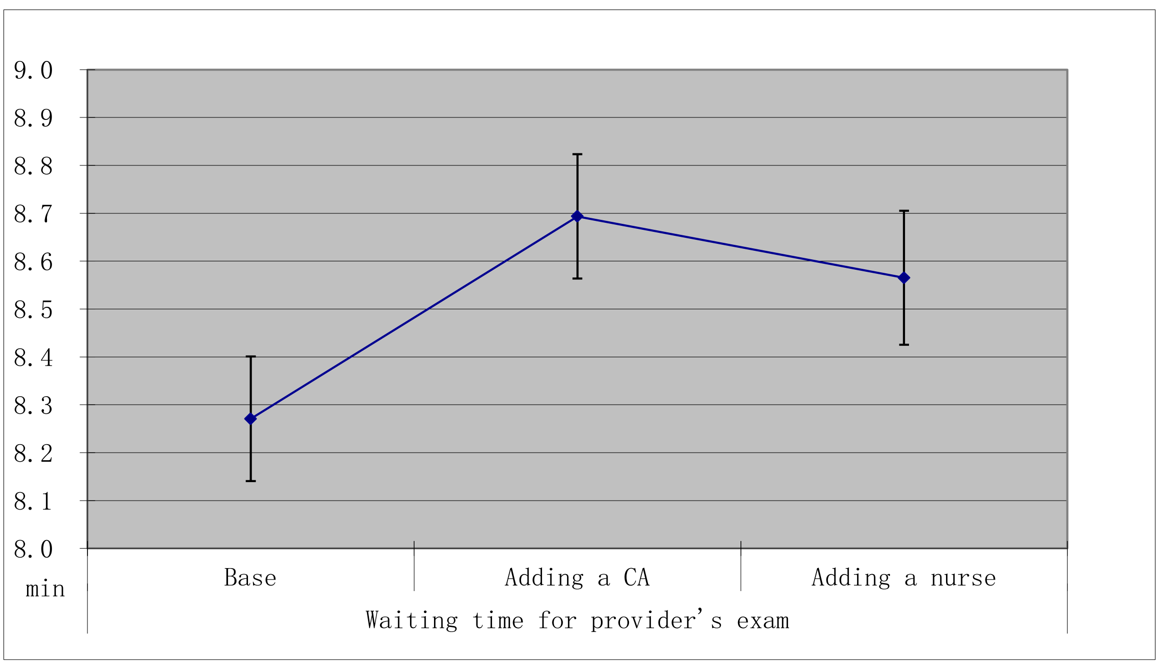


Figure 26: Sensitivity analysis of 3-provider Topfer site (expected waiting time for provider's exam)

For the reasons that we discuss above, adding a nurse or CA does not improve the waiting time for provider's exam, as shown in Figure 26. And adding a nurse outperforms adding a CA in improving the expected waiting time for discharge, as shown in Figure 27. The waiting time for discharge decreases four minutes when an additional nurse is shared in the system.

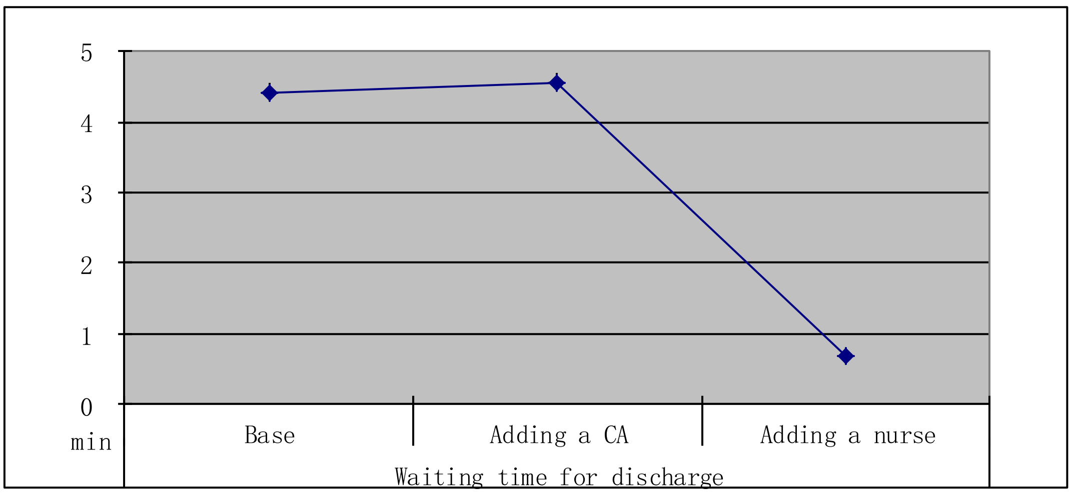
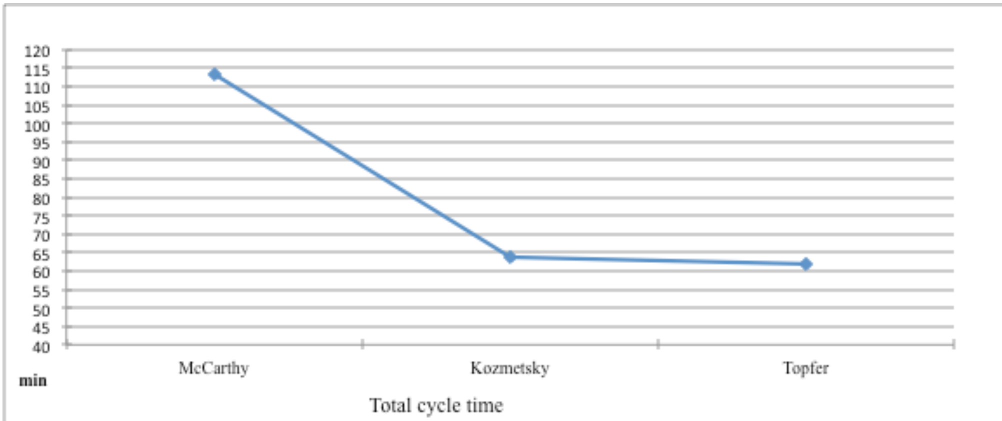


Figure 27: Sensitivity analysis of 3-provider Topfer site (expected waiting time for discharge)

4.4 Comparing performance across three sites

We compare the performance across the three sites using the same baseline scenario (three providers, nine exam rooms, nine nurses and three CAs). The only difference is the input distributions for the service times of the staff of each clinic site, which can be seen in Table 19, Table 21 and Table 25. As shown in Figure 28, when we compare the output results of the three clinic sites, we notice the Topfer site has the best performance among all three sites in expected total cycle time and in waiting time for the three stages, due to the fact that the Topfer site almost always has the lowest input service times at every stage.



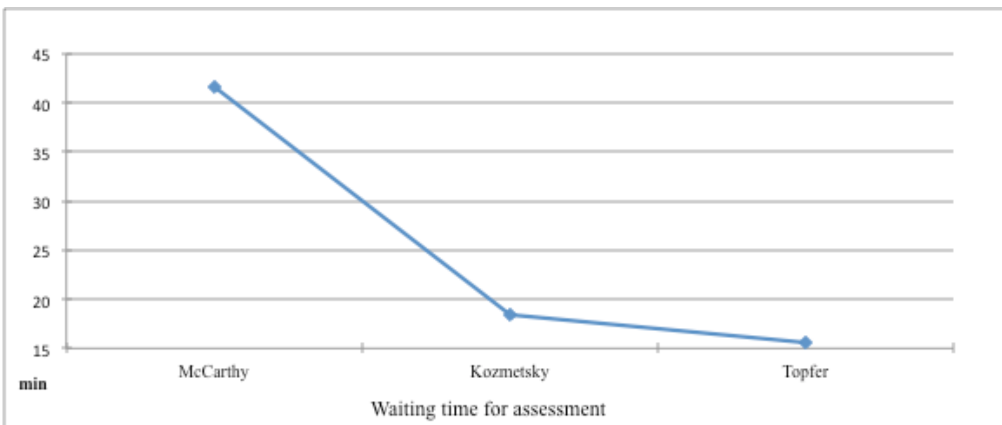
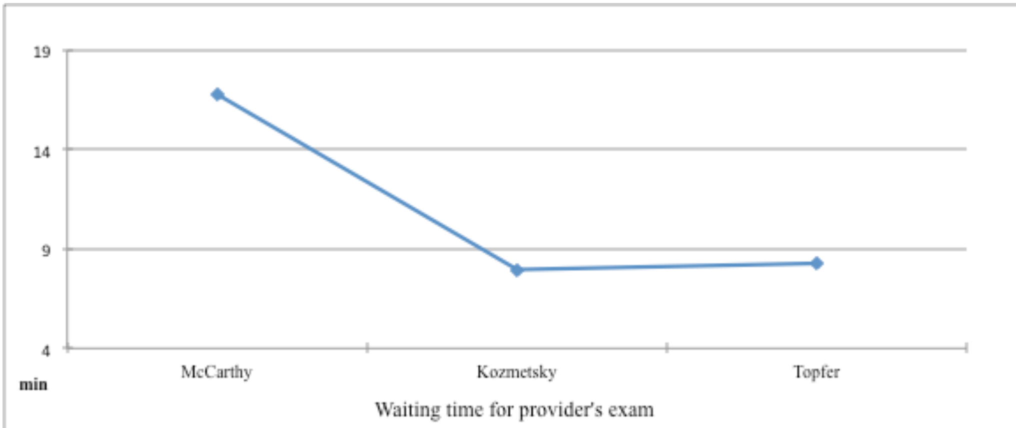


Figure 28: Comparing performance across three sites (While the figure does not depict the confidence intervals, they are the same as shown in the earlier figures in this chapter.)



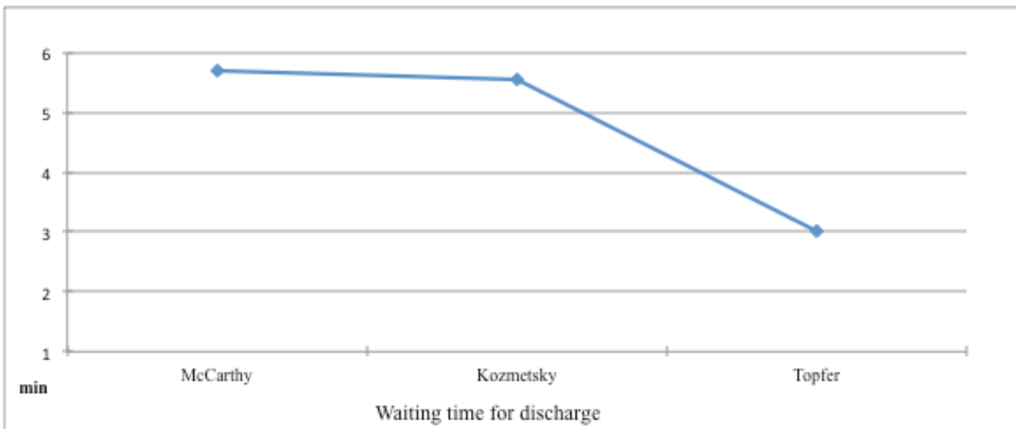


Figure 28 (Continued).

4.5 CONCLUSION AND RECOMMENDATIONS

Comparing the two ways of modeling the arrivals of patients, we found that that the model construct that models the deviation of a patient's arrival time from the appointment time better captures the real system in contrast to ignoring the appointment time and modeling the inter-arrival time of patients. From the simulation results, we find that for all three sites, sharing an additional nurse reduces the expected cycle time significantly, assuming that the service rates at each stage are fixed. If the extra cost for hiring a CA or RN is similar then it is clear that hiring an RN is preferable for improving patient flow. Regardless of cost, the simulation results do not suggest that hiring an additional CA is an effective way to improve patient flow. More generally, this project demonstrates that by using simulation methods, it is possible to improve the patient flow in healthcare service facilities, so that they can better handle challenge of increasing demand in healthcare service.

References:

[1] M. D. McHugh, L.H. Aiken, and R. A. Cooper, The U.S. Presidential election and health care workforce policy, *Policy Politics Nursing Practice*, 9(1), pp. 6-14 , 2008

[2] R. W. Hall, Patient flow, the new queueing theory for healthcare, *OR/MS Today*, 33, pp. 36-40, 2006

[3] M. J. Côté, Understanding patient flow, *Decision Line*, 31(2), pp. 8-9, 2000

[4] Http://www.seton.net/, 22, June, 2011

[5] X. M. Huang, Patient attitude towards waiting in an outpatient clinic and its applications, *Health Services Management Research*, 7(1), pp. 2-8, 1994

[6] A. Wilson and L. Nathan, Understanding benchmarks, *Home Healthcare Nurse,* 21(2), pp. 102-107, 2003

[7] T.W. Nolan, D. M. Berwick, and J. Roessner, *Reducing Delays and Waiting Times Throughout the Healthcare System*, Institute for Healthcare Improvement, 1996

[8] G. J. Langley and C. L. Norman, *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance,* John Wiley and Sons, 2009

[9] A. D. Racine and A. G. Davidson, Use of a time-flow study to improve patient waiting times at an inner-city academic pediatric practice, *Archives of Pediatrics & Adolescent Medicine*, 156(12), pp. 1203-1209, 2002

[10] N. M. Potisek, R. M Malone, B. B. Shilliday, T. J. Ives, P. R Chelminski, D. A. DeWalt, and M. P. Pignone, Use of patient flow analysis to improve patient visit efficiency by decreasing wait time in a primary care-based disease management programs for anticoagulation and chronic pain: A quality improvement study, *BMC Health Service Research* 7(8), 2007

[11] J. Banks and B. L. Nelson, *Discrete-Event Simulation*, Pearson Prentice Hall, 2010

[12] J. F. Merkle, Computer Simulation: A methodology to improve the efficiency in the Brooke Army Medical Center Family Care Clinic (Fellow project), *Journal of Healthcare Management*, 47(1), pp. 58-67, 2002

[13] J. R. Swisher, S. H. Jacobson, J. B. Jun, and O. Balci, Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation, *Computers and Operations Research*, 28(2), pp. 105-125, 2001

[14] J. R. Swisher, B. Jun, S.H. Jacobson, and O. Balci, Simulation of the Queston physician network, *Proceedings of the Winter Simulation Conference,* 1997

[15] T. R. Rohleder, D. P. Bischak, and L. B. Baskin, Modeling patient service centers with simulation and system dynamics, *Health Care Management Science*, 10(1), pp. 1-12, 2007

[16] I. Hill-Smith, Mathematical relationship between waiting times and appointment interval for doctor and patients, *Journal of the Royal College of General Practitioners*, 39(329), pp. 492–494, 1989

[17] F. Dexter, Design of appointment systems for preanesthesia evaluation clinics to minimize patient waiting times: A review of computer simulation and patient survey studies, *Anesthesia & Analgesia*, 89(4), pp. 925-31, 1999